

**Combining cluster sampling and link-tracing  
sampling to estimate the size of a hidden population:  
asymptotic properties of the estimators**

Martín H. Félix Medina

Technical report

Number: FCFM-UAS-2015-01

Class: Research

June 10, 2015

Facultad de Ciencias Físico-Matemáticas

Universidad Autónoma de Sinaloa

Ciudad Universitaria, Culiacán Sinaloa

México

# Combining cluster sampling and link-tracing sampling to estimate the size of a hidden population: asymptotic properties of the estimators

Martín H. Félix Medina\*

Facultad de Ciencias Físico-Matemáticas de la  
Universidad Autónoma de Sinaloa

## Abstract

Félix-Medina and Thompson (2004) proposed a variant of link-tracing sampling to estimate the size of a hidden population such as drug users, sexual workers or homeless people. In their variant a sampling frame of sites where the members of the population tend to gather is constructed. The frame is not assumed to cover the whole population, but only a portion of it. A simple random sample of sites is selected; the people in the sampled sites are identified and are asked to name other members of the population which are added to the sample. Those authors proposed maximum likelihood estimators of the population size which derived from a multinomial model for the numbers of people found in the sampled sites and a model that considers that the probability that a person is named by any element in a particular sampled site (link-probability) does not depend on the named person, that is, that the probabilities are homogeneous. Later, Félix-Medina et al. (2015) proposed unconditional and conditional maximum likelihood estimators of the population size which derived from a model that takes into account the heterogeneity of the link-probabilities. In this work we consider this sampling design and set conditions for a general model for the link-probabilities that guarantee the consistency and asymptotic normality of the estimators of the population size and of the estimators of the parameters of the model for the link-probabilities. In particular we showed that both the unconditional and conditional maximum likelihood estimators of the population size are consistent and have asymptotic normal distributions which are different from each other.

**Key words:** *Asymptotic normality, capture-recapture, chain referral sampling, hard-to-detect population, maximum likelihood estimator, snowball sampling*

---

\*mhfelix@uas.edu.mx

# 1 Introduction

Conventional sampling methods are not appropriate for sampling hidden or hard-to-reach human populations, such as drug users, sexual-workers and homeless people, because of the lack of suitable sampling frames. For this reason, several specific sampling methods for this type of population have been proposed. See Magnani et al. (2005) and Kalton (2009) for reviews of some of them. One of these methods is snowball sampling, also known as link-tracing sampling (LTS) or chain referral sampling. In LTS an initial sample of members of the population is selected and the sample size is increased by asking the people in the initial sample to name other members of the populations. The named people who are not in the initial sample are added to the sample and they are asked to name other members of the population. The sampling process might continue in this way until a stopping rule is satisfied. For reviews of several variants of LTS see Spreen (1992), Thompson and Frank (2000) and Johnston and Sabin (2010).

Félix-Medina and Thompson (2004) proposed a variant of link-tracing sampling (LTS) to estimate the size of a hidden population. In their variant they supposed that a sampling frame of sites where the members of the target population tend to gather can be constructed. As examples of sites are public parks, bars and blocks. It is worth noting that they do not suppose that the frame covers the whole population, but only a portion of it. Then an initial sample of sites is selected by a simple random sampling without replacement design and the members of the population who belong to the sampled sites are identified. Finally the people in the initial sample are asked to name other members of the population and the named persons who are not in the initial sample are included in the sample. Those authors proposed models to describe the number of members of the population who belong to each site in the frame and to describe the probability that a person is linked to a sampled site, that is, that he or she was named by at least one person who belongs to that site. From those models they derived maximum likelihood estimators of the population size. In that work those authors considered that the probability that a person is linked to a site (link-probability) does not depend on the person, but does on the site, that is, they consider homogeneous link-probabilities.

Félix-Medina and Monjardin (2006) considered this same variant of LTS and derived estimators of the population size using a Bayesian-assisted approach, that is, they derived the estimators using the Bayesian approach, but the inferences were made under a frequentist approach. Those authors considered an homogeneous two-stage normal model for the logits of the link-probabilities.

Later Félix-Medina et al. (2015) extended the work by Félix-Medina and Thompson (2004) to the case in which the link-probabilities are heterogeneous, that is, that they depend on the named people. Those authors modeled the heterogeneity of the link-probabilities by means of a mixed logistic normal model proposed by Coull and Agresti (1999) in the context of capture-recapture studies. From this model they derived unconditional and conditional maximum likelihood estimators of the population size.

In this work we consider the variant of the LTS proposed by Félix-Medina and Thompson (2004) and a general model for the link-probabilities from which we derive the forms of the

unconditional and conditional maximum likelihood estimators of the population size. We state conditions that guarantee the consistency and asymptotic normality of both types of estimators, and we proposed estimators of the variances of the estimators of the population size. It is worth noting that our work is based on that by Sanathanan (1972) in which she derived asymptotic properties of both unconditional and conditional maximum likelihood estimators of the size of a multinomial distribution from an incomplete observation of the cell totals which is a situation that occurs in capture-recapture studies. Thus, our work is basically an adaptation of that by Sanathanan (1972) to the estimators used in the sampling variant proposed by Félix-Medina and Thompson (2004).

The structure of this document is the following. In section 2 we describe the variant of LTS proposed by Félix-Medina and Thompson (2004). In section 3 we present probability models that describe the numbers of people that belong to the sites in the frame and the probabilities of links between the members of the population and the sites. From these models we construct the likelihood function that allows us to derive the unconditional and conditional maximum likelihood estimators of the parameters of the assumed model for the link-probabilities and of the population size. In addition, we present conditions that guarantee the consistency of the proposed estimators. In section 4, which is the central part of this paper, we define the asymptotic framework under which are derived the asymptotic properties of the proposed estimators. In section 5 we proposed a method for estimating the variance-covariance matrices of the estimators of the different vectors of parameters that appear in the assumed models. Finally, in section 6 we discuss some points to be considered whenever the results of this paper want to be used in actual situations.

## 2 Link-tracing sampling design

In this section we will describe the LTS variant proposed by Félix-Medina and Thompson (2004). Thus, let  $U$  be a finite population of  $\tau$  people. Let  $U_1$  be the portion of  $U$  that is covered by a sampling frame of  $N$  sites  $A_1, \dots, A_N$ , which are places where members of the population tend to gather. We will assume that each one of the  $\tau_1$  persons who are in  $U_1$  belongs to only one site  $A_i$  in the frame. Notice that this does not imply that a person cannot be found in distinct places, but that, as in ordinary cluster sampling, the researcher has a criterion that allows him or her to assign a person to only one site. Let  $M_i$  be the number of people in  $U_1$  that belong to the site  $A_i$ ,  $i = 1, \dots, N$ . The previous assumption implies that  $\tau_1 = \sum_1^N M_i$ . Let  $\tau_2 = \tau - \tau_1$  be the number of people that belong to the portion  $U_2 = U - U_1$  of  $U$  that is not covered by the sampling frame.

The sampling procedure is as follows. An initial simple random sample without replacement (SRSWOR)  $S_A$  of  $n$  sites  $A_1, \dots, A_n$  is selected from the frame and the members of the population who belong to each sampled site are identified. Let  $S_0$  be the set of people in the initial sample. Notice that the size of  $S_0$  is  $M = \sum_1^n M_i$ . Then from each sampled site  $A_i$ ,  $i = 1, \dots, n$ , the people who belong to that site are asked to name other members of the population. A person and a sampled site are said to be linked if any of the persons who belong to that site names that person. Let  $S_1$  and  $S_2$  be the sets of people in  $U_1 - S_0$  and in

$U_2$ , respectively, who are linked to at least one site in  $S_A$ . Finally, from each named person the following information is obtained: the portion of  $U$  where that person is located, that is,  $U_1 - S_0$ ,  $A_i \in S_A$  or  $U_2$ , and the subset of sampled sites that are linked to him or her.

### 3 Unconditional and conditional maximum likelihood estimators

#### 3.1 Probability models

As in Félix-Medina and Thompson (2004), we will suppose that the numbers  $M_1, \dots, M_N$  of people who belong to the sites  $A_1, \dots, A_N$  are independent Poisson random variables with mean  $\lambda_1$ . Therefore, the joint conditional distribution of  $(M_1, \dots, M_n, \tau_1 - M)$  given that  $\sum_1^N M_i = \tau_1$  is multinomial with probability mass function (pmf):

$$f(m_1, \dots, m_n, \tau_1 - m | \tau_1) = \frac{\tau_1!}{\prod_1^n m_i! (\tau_1 - m)!} \left( \frac{1}{N} \right)^m \left( 1 - \frac{n}{N} \right)^{\tau_1 - m}. \quad (1)$$

To model the links between the members of the population and the sampled sites we will define for person  $j$  in  $U_k - S_0$  the vector of link-indicator variables  $\mathbf{X}_j^{(k)} = (X_{1j}^{(k)}, \dots, X_{nj}^{(k)})$ , where  $X_{ij}^{(k)} = 1$  if person  $j$  is linked to site  $A_i$  and  $X_{ij}^{(k)} = 0$  otherwise. Notice that  $\mathbf{X}_j^{(k)}$  indicates which sites in  $S_A$  are linked to person  $j$ . We will suppose that given  $S_A$ , and consequently the values  $M_i$ s of the sampled sites, the  $X_{ij}^{(k)}$ s are Bernoulli random variables with means  $p_{ij}^{(k)}$ s and that the vectors  $\mathbf{X}_j^{(k)}$  are independent. Let  $\Omega = \{(x_1, \dots, x_n) : x_i = 0, 1; i = 1, \dots, n\}$ , that is, the set of all the  $n$ -dimensional vectors such that each one of their elements is 0 or 1. For  $\mathbf{x} = (x_1, \dots, x_n) \in \Omega$  we will denote by  $\pi_{\mathbf{x}}^{(k)}$  the probability that the vector of link-indicator variables associated with a randomly selected person from  $U_k - S_0$  equals  $\mathbf{x}$ , that is, the probability that the person is linked only to the sites  $A_i$  such that the  $i$ -th element  $x_i$  of  $\mathbf{x}$  equals 1. We will suppose that  $\pi_{\mathbf{x}}^{(k)}$  depends on a  $q_k$ -dimensional parameter  $\boldsymbol{\theta}_k = (\theta_1^{(k)}, \dots, \theta_{q_k}^{(k)}) \in \boldsymbol{\Theta}_k \subseteq \mathbb{R}^{q_k}$ , that is,  $\pi_{\mathbf{x}}^{(k)} = \pi_{\mathbf{x}}^{(k)}(\boldsymbol{\theta}_k)$ ,  $k = 1, 2$ . In this work we will assume that  $\boldsymbol{\theta}_k$  does not depend on the observed  $M_i$ s.

Similarly, for person  $j$  in  $A_i \in S_A$ , we will define the vector of link-indicator variables  $\mathbf{X}_j^{(A_i)} = (X_{1j}^{(A_i)}, \dots, X_{i-1j}^{(A_i)}, X_{i+1j}^{(A_i)}, \dots, X_{nj}^{(A_i)})$ , where  $X_{i'j}^{(A_i)} = 1$  if person  $j$  is linked to site  $A_{i'}$ ,  $i' = 1, \dots, n$ ,  $i' \neq i$  and  $X_{i'j}^{(A_i)} = 0$  otherwise. We will suppose that given  $S_A$  the  $X_{i'j}^{(A_i)}$ s are Bernoulli random variables with means  $p_{i'j}^{(A_i)}$ s and that the vectors  $\mathbf{X}_j^{(A_i)}$  are independent. For each  $A_i \in S_A$ , let  $\Omega_{-i} = \{(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) : x_{i'} = 0, 1; i' \neq i, i' = 1, \dots, n\}$ , that is, the set of all  $(n - 1)$ -dimensional vectors obtained from the vectors in  $\Omega$  by omitting their  $i$ -th coordinate. For  $\mathbf{x} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \in \Omega_{-i}$  we will denote by  $\pi_{\mathbf{x}}^{(A_i)}$  the probability that the vector of link-indicator variables associated with a randomly selected person from  $A_i$  equals  $\mathbf{x}$ . We will suppose that  $\pi_{\mathbf{x}}^{(A_i)}$  depends on the  $q_1$ -dimensional parameter  $\boldsymbol{\theta}_1 = (\theta_1^{(1)}, \dots, \theta_{q_1}^{(1)}) \in \boldsymbol{\Theta}_1$ , that is,  $\pi_{\mathbf{x}}^{(A_i)} = \pi_{\mathbf{x}}^{(A_i)}(\boldsymbol{\theta}_1)$ ,  $i = 1, \dots, n$ .

For instance, Félix-Medina and Monjardin (2006) modeled the link-probability between person  $j$  in  $U_k - A_i$  and site  $A_i \in S_A$  by  $p_{ij}^{(k)} = \Pr(X_{ij}^{(k)} = 1 | S_A) = \exp(\alpha_i^{(k)}) / [1 + \exp(\alpha_i^{(k)})]$ , where the conditional distribution of  $\alpha_i^{(k)}$  given  $\psi_k$  is normal with mean  $\psi_k$  and variance  $\sigma_k^2$ , which we denote by  $\alpha_i^{(k)} | \psi_k \sim N(\psi_k, \sigma_k^2)$  and  $\psi_k \sim N(\mu_k, \gamma_k^2)$ . Thus, in this case  $\theta_k = (\mu_k, \gamma_k, \sigma_k) \in \Theta_k = \mathbb{R} \times (0, \infty) \times (0, \infty)$ , and

$$\pi_{\mathbf{x}}^{(k)}(\theta_k) = \left[ \int \int \frac{\exp(\alpha)}{1 + \exp(\alpha)} f_k(\alpha | \psi) f_k(\psi) d\alpha d\psi \right]^t \times \left[ \int \int \frac{1}{1 + \exp(\alpha)} f_k(\alpha | \psi) f_k(\psi) d\alpha d\psi \right]^{n-t},$$

where  $\mathbf{x} = (x_1, \dots, x_n) \in \Omega$ ,  $t = \sum_1^n x_i$ , and  $f_k(\alpha | \psi)$  and  $f_k(\psi)$  denote the probability density functions of the distributions  $N(\psi_k, \sigma_k^2)$  and  $N(\mu_k, \gamma_k^2)$ , respectively. It is worth noting that those authors did not compute  $\pi_{\mathbf{x}}^{(k)}(\theta_k)$  because they followed a Bayesian approach and focused on computing the posterior distribution of the parameters.

As another example, Félix-Medina et al. (2015) modeled the link-probability between person  $j$  in  $U_k - A_i$  and site  $A_i \in S_A$  by the following Rasch model:  $p_{ij}^{(k)} = \Pr(X_{ij}^{(k)} = 1 | S_A) = \exp(\alpha_i^{(k)} + \beta_j^{(k)}) / [1 + \exp(\alpha_i^{(k)} + \beta_j^{(k)})]$ , where  $\alpha_i^{(k)}$  is a fixed (not random) effect associated with the site  $A_i$  and  $\beta_j^{(k)}$  is a normal random effect with mean zero and variance  $\sigma_k^2$  associated with person  $j$  in  $U_k - A_i$ . Therefore

$$\pi_{\mathbf{x}}^{(k)}(\theta_k) = \int \prod_{i=1}^n \frac{\exp[x_i (\alpha_i^{(k)} + \sigma_k z)]}{1 + \exp(\alpha_i^{(k)} + \sigma_k z)} \phi(z) dz,$$

where  $\mathbf{x} = (x_1, \dots, x_n) \in \Omega$ ,  $\theta_k = (\alpha_1^{(k)}, \dots, \alpha_n^{(k)}, \sigma_k) \in \Theta_k = \mathbb{R}^n \times (0, \infty)$  and  $\phi(\cdot)$  denotes the probability density function of the standard normal distribution. Those authors compute  $\pi_{\mathbf{x}}^{(k)}(\theta_k)$  by means of Gaussian quadrature formula.

Notice that in the first example the parameter  $\theta_k$  is defined previously to the selection of the initial sample because the  $\alpha_i^{(k)}$ s are a random sample from a probability distribution indexed by  $\theta_k$  and consequently this parameter does not represent characteristics of the particular selected sample. On the other hand, in the second example the parameter  $\theta_k$  is defined once the initial sample of sites is selected because the  $\alpha_i^{(k)}$ s represent characteristics of the particular sites in  $S_A$ . Therefore, as long as  $\theta_k$  does not depend on the  $M_i$ s the results derived in this work are valid for both cases.

### 3.2 Likelihood function

To compute the likelihood function we will factorize it into different components. One component,  $L_{MULT}(\tau_1)$ , is given by the probability of observing the particular sizes  $m_1, \dots, m_n$  of the sites in  $S_A$ ; therefore, it is specified by the multinomial distribution (1). Two additional

factors are given by the probabilities of the configurations of the links between the people in  $U_k - S_0$ ,  $k = 1, 2$ , and the sites  $A_i \in S_A$ . To obtain those factors we will denote by  $R_{\mathbf{x}}^{(k)}$ ,  $\mathbf{x} = (x_1, \dots, x_n) \in \Omega$ , the random variable that indicates the number of distinct people in  $U_k - S_0$  whose vectors of link-indicator variables are equal to  $\mathbf{x}$ , and by  $R_k$  the random variable that indicates the number of distinct people in  $U_k - S_0$  who are linked to at least one site  $A_i \in S_A$ . Notice that  $R_k = \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} R_{\mathbf{x}}^{(k)}$ , where  $\mathbf{0}$  denotes the  $n$ -dimensional vector of zeros.

Because of the assumptions we made about the vectors  $\mathbf{X}_j^{(k)}$  of link-indicator variables we have that the conditional joint probability distribution of the variables  $\{R_{\mathbf{x}}^{(1)}\}_{\mathbf{x} \in \Omega}$  given  $S_A$  is a multinomial distribution with parameter of size  $\tau_1 - m$  and probabilities  $\{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1)\}_{\mathbf{x} \in \Omega}$ , whereas that of the variables  $\{R_{\mathbf{x}}^{(2)}\}_{\mathbf{x} \in \Omega}$  is a multinomial distribution with parameter of size  $\tau_2$  and probabilities  $\{\pi_{\mathbf{x}}^{(2)}(\boldsymbol{\theta}_2)\}_{\mathbf{x} \in \Omega}$ . Therefore, the factors of the likelihood function associated with the probabilities of the configurations of links between the people in  $U_k - S_0$ ,  $k = 1, 2$ , and the sites  $A_i \in S_A$  are

$$L_1(\tau_1, \boldsymbol{\theta}_1) = \frac{(\tau_1 - m)!}{(\tau_1 - m - r_1)! \prod_{\mathbf{x} \in \Omega} r_{\mathbf{x}}^{(1)}!} \prod_{\mathbf{x} \in \Omega} [\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1)]^{r_{\mathbf{x}}^{(1)}} \quad (2)$$

and

$$L_2(\tau_2, \boldsymbol{\theta}_2) = \frac{\tau_2!}{(\tau_2 - r_2)! \prod_{\mathbf{x} \in \Omega} r_{\mathbf{x}}^{(2)}!} \prod_{\mathbf{x} \in \Omega} [\pi_{\mathbf{x}}^{(2)}(\boldsymbol{\theta}_2)]^{r_{\mathbf{x}}^{(2)}}.$$

Notice that  $r_0^{(1)} = \tau_1 - m - r_1$  and  $r_0^{(2)} = \tau_2 - r_2$ .

The last factor of the likelihood function is given by the probability of the configuration of links between the people in  $S_0$  and the sites  $A_i \in S_A$ . To obtain this factor, we will denote by  $R_{\mathbf{x}}^{(A_i)}$ ,  $\mathbf{x} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \in \Omega_{-i}$ , the random variable that indicates the number of distinct people in  $A_i \in S_A$  such that their vectors of link-indicator variables equal  $\mathbf{x}$  and by  $R^{(A_i)}$  the random variable that indicates the number of distinct people in  $A_i \in S_A$  who are linked to at least one site  $A_j \in S_A$ ,  $j \neq i$ . Notice that  $R^{(A_i)} = \sum_{\mathbf{x} \in \Omega_{-i} - \{\mathbf{0}\}} R_{\mathbf{x}}^{(A_i)}$ , where  $\mathbf{0}$  denotes the  $(n-1)$ -dimensional vector of zeros and  $R_0^{(A_i)} = m_i - R^{(A_i)}$ . Then, as in the previous cases, the conditional joint probability distribution of the variables  $\{R_{\mathbf{x}}^{(A_i)}\}_{\mathbf{x} \in \Omega_{-i}}$  given  $S_A$  is a multinomial distribution with parameter of size  $m_i$  and probabilities  $\{\pi_{\mathbf{x}}^{(A_i)}(\boldsymbol{\theta}_1)\}_{\mathbf{x} \in \Omega_{-i}}$ . Therefore, the probability of the configuration of links between the people in  $S_0$  and the sites  $A_i \in S_A$  is given by the product of the previous multinomial probabilities (one for each  $A_i \in S_A$ ), and consequently the factor of the likelihood function associated with that probability is

$$L_0(\boldsymbol{\theta}_1) = \prod_{i=1}^n \frac{m_i!}{\prod_{\mathbf{x} \in \Omega} r_{\mathbf{x}}^{(A_i)}!} \prod_{\mathbf{x} \in \Omega_{-i}} [\pi_{\mathbf{x}}^{(A_i)}(\boldsymbol{\theta}_1)]^{r_{\mathbf{x}}^{(A_i)}} [\pi_{\mathbf{0}}^{(A_i)}(\boldsymbol{\theta}_1)]^{m_i - r^{(A_i)}}.$$

From the previous results we have that the maximum likelihood function is given by

$$L(\tau_1, \tau_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = L_{(1)}(\tau_1, \boldsymbol{\theta}_1) L_{(2)}(\tau_2, \boldsymbol{\theta}_2),$$

where

$$\begin{aligned} L_{(1)}(\tau_1, \boldsymbol{\theta}_1) &= L_{MULT}(\tau_1) L_1(\tau_1, \boldsymbol{\theta}_1) L_0(\boldsymbol{\theta}_1) \quad \text{and} \\ L_{(2)}(\tau_2, \boldsymbol{\theta}_2) &= L_2(\tau_2, \boldsymbol{\theta}_2). \end{aligned} \quad (3)$$

### 3.3 Unconditional and conditional maximum likelihood estimators of $(\tau_k, \boldsymbol{\theta}_k^*)$

In this section we will derive unconditional and conditional maximum likelihood estimators of the parameters of the previously specified models. Henceforth we will suppose that conditional on the initial sample  $S_A$  of sites the following “regularity conditions” are satisfied:

- (1)  $\boldsymbol{\theta}_k^*$  is the true value of  $\boldsymbol{\theta}_k$ .
- (2)  $\boldsymbol{\theta}_k^*$  is an interior point of  $\Theta_k$ .
- (3)  $\pi_{\mathbf{x}}^{(k)}(\boldsymbol{\theta}_k^*) > 0$ ,  $\mathbf{x} \in \Omega$  and  $\pi_{\mathbf{x}}^{(A_i)}(\boldsymbol{\theta}_1^*) > 0$ ,  $\mathbf{x} \in \Omega_{-i}$ ,  $i = 1, \dots, n$ .
- (4)  $\partial \pi_{\mathbf{x}}^{(k)}(\boldsymbol{\theta}_k)/\partial \theta_j^{(k)}$ ,  $\mathbf{x} \in \Omega$  and  $\partial \pi_{\mathbf{x}}^{(A_i)}(\boldsymbol{\theta}_1)/\partial \theta_j^{(1)}$ ,  $\mathbf{x} \in \Omega_{-i}$ ,  $i = 1, \dots, n$ ;  $j = 1, \dots, q_k$ , exist at any  $\boldsymbol{\theta}_k \in \Theta_k$  and  $\boldsymbol{\theta}_1 \in \Theta_1$ , and are continuous in neighborhoods of  $\boldsymbol{\theta}_k^*$  and  $\boldsymbol{\theta}_1^*$ , respectively.
- (5) Given a  $\delta_1 > 0$ , it is possible to find an  $\varepsilon_1 > 0$  such that

$$\begin{aligned} & \inf_{\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_1^*\| > \delta_1} \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)} \ln \left\{ \frac{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*) / [1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)]}{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1) / [1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1)]} \right\} \\ & + \frac{1}{(N - n) [1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)]} \sum_{i=1}^n \sum_{\mathbf{x} \in \Omega_{-i}} \pi_{\mathbf{x}}^{(A_i)}(\boldsymbol{\theta}_1^*) \ln \left[ \frac{\pi_{\mathbf{x}}^{(A_i)}(\boldsymbol{\theta}_1^*)}{\pi_{\mathbf{x}}^{(A_i)}(\boldsymbol{\theta}_1)} \right] \geq \varepsilon_1. \end{aligned}$$

- (6) Given a  $\delta_2 > 0$ , it is possible to find an  $\varepsilon_2 > 0$  such that

$$\inf_{\|\boldsymbol{\theta}_2 - \boldsymbol{\theta}_2^*\| > \delta_2} \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{\pi_{\mathbf{x}}^{(2)}(\boldsymbol{\theta}_2^*)}{1 - \pi_{\mathbf{0}}^{(2)}(\boldsymbol{\theta}_2^*)} \ln \left\{ \frac{\pi_{\mathbf{x}}^{(2)}(\boldsymbol{\theta}_2^*) / [1 - \pi_{\mathbf{0}}^{(2)}(\boldsymbol{\theta}_2^*)]}{\pi_{\mathbf{x}}^{(2)}(\boldsymbol{\theta}_2) / [1 - \pi_{\mathbf{0}}^{(2)}(\boldsymbol{\theta}_2)]} \right\} \geq \varepsilon_2.$$

**Remark 1.** For a differentiable function  $f : \mathbb{R}^q \rightarrow \mathbb{R}$ , the notation  $\partial f(\mathbf{x}_0)/\partial x_j$  represents  $\partial f(\mathbf{x})/\partial x_j|_{\mathbf{x}=\mathbf{x}_0}$ .



The regularity conditions (1)-(4) and (6) or conditions equivalent to them have been assumed by several authors such as Birch (1964), Rao (1973, Ch. 5), Bishop et al. (1975, Ch. 14), Sanathanan (1972) and Agresti (2002, Ch. 14), among others, in the context of deriving asymptotic properties of estimators of the parameters of models for the probabilities of a multinomial distribution. The particular form of condition (6) comes from Sanathanan (1972) who took it from the first edition of Rao (1973, Ch. 5) and it is known as a *strong identifiability* condition. Condition (5) is a modification of (6) to meet the requirements of our particular sampling design. In general, these conditions imply the existence and consistency of the UMLEs and CMLEs of  $\theta_1^*$  and  $\theta_2^*$ , and that they can be obtained deriving the likelihood function with respect to  $\theta_1$  and  $\theta_2$ .

### 3.3.1 Unconditional and conditional maximum likelihood estimators of $\tau_1$ and $\theta_1^*$

Let us firstly consider the unconditional maximum likelihood estimators (UMLEs)  $\hat{\tau}_1^{(U)}$  and  $\hat{\theta}_1^{(U)}$  of  $\tau_1$  and  $\theta_1^*$ . The log-likelihood function of  $\tau_1$  and  $\theta_1$  is

$$\begin{aligned} l_{(1)}(\tau_1, \theta_1) &= \ln[L_{(1)}(\tau_1, \theta_1)] \\ &= \ln(\tau_1!) - \ln[(\tau_1 - m - r_1)!] + \tau_1 \ln(1 - n/N) \\ &\quad + \sum_{\mathbf{x} \in \Omega} r_{\mathbf{x}}^{(1)} \ln [\pi_{\mathbf{x}}^{(1)}(\theta_1)] + \sum_{i=1}^n \sum_{\mathbf{x} \in \Omega_{-i}} r_{\mathbf{x}}^{(A_i)} \ln [\pi_{\mathbf{x}}^{(A_i)}(\theta_1)] + C, \end{aligned}$$

where  $C$  does not depend on  $\tau_1$  and  $\theta_1$ , and recall that  $r_0^{(1)} = \tau_1 - m - r_1$  and  $r_0^{(A_i)} = m_i - r^{(A_i)}$ . Then, the UMLE  $\hat{\theta}_1^{(U)}$  of  $\theta_1^*$  is the solution to the following equations:

$$\frac{\partial l_{(1)}(\tau_1, \theta_1)}{\partial \theta_j^{(1)}} = \sum_{\mathbf{x} \in \Omega} \frac{r_{\mathbf{x}}^{(1)}}{\pi_{\mathbf{x}}^{(1)}(\theta_1)} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\theta_1)}{\partial \theta_j^{(1)}} + \sum_{i=1}^n \sum_{\mathbf{x} \in \Omega_{-i}} \frac{r_{\mathbf{x}}^{(A_i)}}{\pi_{\mathbf{x}}^{(A_i)}(\theta_1)} \frac{\partial \pi_{\mathbf{x}}^{(A_i)}(\theta_1)}{\partial \theta_j^{(1)}} = 0, \quad j = 1, \dots, q_1. \quad (4)$$

Since  $\tau_1$  is an integer we will use the “ratio method” to maximize  $L_{(1)}(\tau_1, \theta_1)$ . [See Feller (1968, Ch. 3).] Thus

$$\frac{L_{(1)}(\tau_1, \theta_1)}{L_{(1)}(\tau_1 - 1, \theta_1)} = \frac{\tau_1(1 - n/N)\pi_0^{(1)}(\theta_1)}{(\tau_1 - m - r_1)}.$$

Since this ratio is greater than or equal to 1 if  $\tau_1 \leq (m + r_1) / [1 - (1 - n/N)\pi_0^{(1)}(\theta_1)]$  and it is smaller than or equal to 1 if  $\tau_1$  is greater than or equal to that quantity, it follows that  $\hat{\tau}_1^{(U)}$  is given by

$$\hat{\tau}_1^{(U)} = \left\lfloor \frac{M + R_1}{1 - (1 - n/N)\pi_0^{(1)}(\hat{\theta}_1^{(U)})} \right\rfloor, \quad (5)$$

where  $\lfloor x \rfloor$  denotes the largest integer not greater than  $x$ . Notice that the right hand-side of (5) is not a closed form for  $\hat{\tau}_1^{(U)}$  since this expression depends on  $\hat{\theta}_1^{(U)}$ . In fact,  $\hat{\tau}_1^{(U)}$  and  $\hat{\theta}_1^{(U)}$  are

obtained by simultaneously solving the set of equations (4) and (5), which is generally done by numerical methods.

Let us now consider the conditional maximum likelihood estimators (CMLEs)  $\hat{\tau}_1^{(C)}$  and  $\hat{\theta}_1^{(C)}$  of  $\tau_1$  and  $\theta_1^*$ . It is worth noting that this type of estimators was proposed by Sanathanan (1972) in the context of estimating the parameter of size of a multinomial distribution from an incomplete observation of the cell frequencies. The approach we will follow to derive  $\hat{\tau}_1^{(C)}$  and  $\hat{\theta}_1^{(C)}$  is an adaptation of Sanathanan's (1972) approach to our case. Thus, from (2) we have that

$$\begin{aligned}
L_1(\tau_1, \theta_1) &= f(\{r_{\mathbf{x}}^{(1)}\}_{\mathbf{x} \in \Omega} | \{m_i\}, \tau_1, \theta_1) \\
&= f(\{r_{\mathbf{x}}^{(1)}\}_{\mathbf{x} \in \Omega - \{0\}} | r_1, \{m_i\}, \tau_1, \theta_1) f(r_1 | \{m_i\}, \tau_1, \theta_1) \\
&= \frac{r_1!}{\prod_{\mathbf{x} \in \Omega - \{0\}} r_{\mathbf{x}}^{(1)}!} \prod_{\mathbf{x} \in \Omega - \{0\}} \left[ \frac{\pi_{\mathbf{x}}^{(1)}(\theta_1)}{1 - \pi_0^{(1)}(\theta_1)} \right]^{r_{\mathbf{x}}^{(1)}} \\
&\quad \times \frac{(\tau_1 - m)!}{(\tau_1 - m - r_1)! r_1!} \left[ 1 - \pi_0^{(1)}(\theta_1) \right]^{r_1} \left[ \pi_0^{(1)}(\theta_1) \right]^{\tau_1 - m - r_1} \\
&= L_{11}(\theta_1) L_{12}(\tau_1, \theta_1)
\end{aligned} \tag{6}$$

Notice that the first factor  $L_{11}(\theta_1)$  is given by the joint pmf of the multinomial distribution with parameter of size  $r_1$  and probabilities  $\left\{ \pi_{\mathbf{x}}^{(1)}(\theta_1) / \left[ 1 - \pi_0^{(1)}(\theta_1) \right] \right\}_{\mathbf{x} \in \Omega - \{0\}}$  and that this distribution does not depend on  $\tau_1$ . Note also that the second factor  $L_{12}(\tau_1, \theta_1)$  is given by the pmf of the binomial distribution with parameter of size  $\tau_1 - m$  and probability  $1 - \pi_0^{(1)}(\theta_1)$ . Thus, the CMLE  $\hat{\theta}_1^{(C)}$  of  $\theta_1^*$  is the solution to the following system of equations:

$$\begin{aligned}
\frac{\partial}{\partial \theta_j^{(1)}} \ln[L_{11}(\theta_1) L_{12}(\theta_1)] &= \sum_{\mathbf{x} \in \Omega - \{0\}} \frac{r_{\mathbf{x}}^{(1)}}{\pi_{\mathbf{x}}^{(1)}(\theta_1)} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\theta_1)}{\partial \theta_j^{(1)}} + \frac{r_1}{1 - \pi_0^{(1)}(\theta_1)} \frac{\partial \pi_0^{(1)}(\theta_1)}{\partial \theta_j^{(1)}} \\
&\quad + \sum_{i=1}^n \sum_{\mathbf{x} \in \Omega - i} \frac{r_{\mathbf{x}}^{(A_i)}}{\pi_{\mathbf{x}}^{(A_i)}(\theta_1)} \frac{\partial \pi_{\mathbf{x}}^{(A_i)}(\theta_1)}{\partial \theta_j^{(1)}} = 0, \quad j = 1, \dots, q_1.
\end{aligned} \tag{7}$$

The CMLE  $\hat{\tau}_1^{(C)}$  of  $\tau_1$  is obtained by the ratio method. Thus, since

$$\frac{L_{MULT}(\tau_1) L_{12}(\tau_1, \theta_1)}{L_{MULT}(\tau_1 - 1) L_{12}(\tau_1 - 1, \theta_1)} = \frac{\tau_1 (1 - n/N) \pi_0^{(1)}(\theta_1)}{(\tau_1 - m - r_1)},$$

it follows that

$$\hat{\tau}_1^{(C)} = \left\lceil \frac{M + R_1}{1 - (1 - n/N) \pi_0^{(1)}(\hat{\theta}_1^{(C)})} \right\rceil. \tag{8}$$

Note that (8) is a closed form for  $\hat{\tau}_1^{(C)}$  since  $\hat{\theta}_1^{(C)}$  is firstly obtained from (7).

### 3.3.2 Unconditional and conditional maximum likelihood estimators of $\tau_2$ and $\theta_2^*$

By a similar analysis as that conducted in the previous subsection we have that the UMLEs  $\hat{\tau}_2^{(U)}$  and  $\hat{\theta}_2^{(U)}$  of  $\tau_2$  and  $\theta_2^*$  are the solution to the following equations:

$$\sum_{\mathbf{x} \in \Omega} \frac{r_{\mathbf{x}}^{(2)}}{\pi_{\mathbf{x}}^{(2)}(\theta_2)} \frac{\partial \pi_{\mathbf{x}}^{(2)}(\theta_2)}{\partial \theta_j^{(2)}} = 0, \quad j = 1, \dots, q_2$$

and

$$\hat{\tau}_2^{(U)} = \left[ \frac{R_2}{1 - \pi_0^{(2)}(\hat{\theta}_2^{(U)})} \right]. \quad (9)$$

where recall that  $r_0^{(2)} = \tau_2 - r_2$ .

With respect to the conditional estimators, we have that the CMLE  $\hat{\theta}_2^{(C)}$  of  $\theta_2^*$  is the solution to the following equations:

$$\sum_{\mathbf{x} \in \Omega - \{0\}} \frac{r_{\mathbf{x}}^{(2)}}{\pi_{\mathbf{x}}^{(2)}(\theta_2)} \frac{\partial \pi_{\mathbf{x}}^{(2)}(\theta_2)}{\partial \theta_j^{(2)}} + \frac{r_2}{1 - \pi_0^{(2)}(\theta_2)} \frac{\partial \pi_0^{(2)}(\theta_2)}{\partial \theta_j^{(2)}} = 0, \quad j = 1, \dots, q_2.$$

The CMLE  $\hat{\tau}_2^{(C)}$  of  $\tau_2$  is given by (9), but replacing  $\hat{\theta}_2^{(U)}$  by  $\hat{\theta}_2^{(C)}$ . Note that in this case (9) is a closed form for  $\hat{\tau}_2^{(C)}$ .

### 3.3.3 Unconditional and conditional maximum likelihood estimators of $\tau = \tau_1 + \tau_2$

The UMLE and CMLE of  $\tau = \tau_1 + \tau_2$  are given by  $\hat{\tau}^{(U)} = \hat{\tau}_1^{(U)} + \hat{\tau}_2^{(U)}$  and  $\hat{\tau}^{(C)} = \hat{\tau}_1^{(C)} + \hat{\tau}_2^{(C)}$ , respectively.

## 4 Asymptotic properties of the unconditional and conditional maximum likelihood estimators

The structure of this section is as follows. Firstly we will define the asymptotic framework under which we will derive the asymptotic properties of the estimators. Next we will state and prove a theorem that guarantees the asymptotic multivariate normal distribution of any estimator of  $(\tau_1, \theta_1^*)$  that satisfies the conditions expressed in the theorem. Since not any estimator of  $(\tau_1, \theta_1^*)$  satisfies the conditions of the theorem, in particular the CMLE does not, we will state and prove another theorem that guarantees the asymptotic multivariate normal distribution of any estimator of  $\theta_1^*$  that satisfies the conditions of that theorem. Then, we will prove that the UMLE of  $(\tau_1, \theta_1^*)$  satisfies the conditions of the first theorem, whereas the CMLE of  $\theta_1^*$  satisfies those of the second one. In addition, we will prove that in spite of that result, the CMLE  $\hat{\tau}_1^{(C)}$  does have an asymptotic normal distribution although it is not the same as that of  $\hat{\tau}_1^{(U)}$ . After that we will consider the asymptotic properties of estimators of

$(\tau_2, \theta_2^*)$ . Since this problem is exactly the same as that considered by Sanathanan (1972), we will only state a theorem that guarantees the asymptotic multivariate normal distribution of any estimator of  $(\tau_2, \theta_2^*)$  that satisfies the conditions expressed in the theorem, but we will omit its proof, as well as the proofs that both the UMLE and the CMLE of  $(\tau_2, \theta_2^*)$  satisfy the conditions of that theorem. Finally, we will obtain the asymptotic properties of the estimators  $\hat{\tau}^{(U)}$  and  $\hat{\tau}^{(C)}$  of  $\tau$ .

## 4.1 Basic assumptions

To derive the asymptotic properties of the UMLEs and CMLEs of  $\tau_k$  and  $\theta_k^*$ ,  $k = 1, 2$ , we will make the following assumptions:

- A.  $\tau_k \rightarrow \infty$ ,  $k = 1, 2$ .
- B.  $\tau_k/\tau \rightarrow \alpha_k$ ,  $0 < \alpha_k < 1$ ,  $k = 1, 2$ .
- C.  $N$  and  $n$  are fixed positive integer numbers.

For convenience of notation, we will put  $\tau_k$  either as a subscript or a superscript of every term that depends on  $\tau_k$ ,  $k = 1, 2$ . In addition, convergence in distribution will be denoted by  $\xrightarrow{D}$  and convergence in probability by  $\xrightarrow{P}$ .

Notice that from (1) it follows that the conditional distribution of  $M_i^{(\tau_1)}$  given  $\tau_1$  is binomial with parameter of size  $\tau_1$  and probability  $1/N$ , that is  $M_i^{(\tau_1)}|\tau_1 \sim \text{Bin}(\tau_1, 1/N)$ ; consequently  $M_i^{(\tau_1)}/\tau_1$  is stochastically bounded, that is,  $M_i^{(\tau_1)} = O_p(\tau_1)$ . This means that the size of  $U_1^{(\tau_1)}$  is increased by increasing the sizes of the clusters, even though their number  $N$  is kept fixed. In the same manner, the number of people in the initial sample  $S_0^{(\tau_1)}$ , given by  $M^{(\tau_1)} = \sum_1^n M_i^{(\tau_1)}|\tau_1 \sim \text{Bin}(\tau_1, n/N)$ , is increased because of the increasing of  $M_i^{(\tau_1)}$ ,  $i = 1, \dots, n$ , even though  $n$  is kept fixed. On the other hand, since  $\tau_1 - M^{(\tau_1)}|\tau_1 \sim \text{Bin}(\tau_1, 1 - n/N)$ ,  $R_1^{(\tau_1)}|S_A^{(\tau_1)} \sim \text{Bin}(\tau_1 - M^{(\tau_1)}, 1 - \pi_0^{(1)})$  and  $R_2^{(\tau_2)}|S_A^{(\tau_1)} \sim \text{Bin}(\tau_2, 1 - \pi_0^{(2)})$ , it follows that  $R_1^{(\tau_1)}|\tau_1 \sim \text{Bin}\left[\tau_1, (1 - n/N) \left(1 - \pi_0^{(1)}\right)\right]$  and  $R_2^{(\tau_2)}|\tau_2 \sim \text{Bin}(\tau_2, 1 - \pi_0^{(2)})$ ; therefore  $R_1^{(\tau_1)} = O_p(\tau_1)$  and  $R_2^{(\tau_2)} = O_p(\tau_2)$ . Thus, the sizes of the sets  $S_1^{(\tau_1)}$  and  $S_2^{(\tau_2)}$  are increased because  $\tau_1$  and  $\tau_2$  are increased even though the probabilities  $\{\pi_x^{(1)}\}_{x \in \Omega}$  and  $\{\pi_x^{(2)}\}_{x \in \Omega}$  are kept fixed.

We will end this subsection presenting the conditional and unconditional distributions of the variables  $R_x^{(\tau_1)}$ ,  $R_x^{(A_i)}$  and  $R_x^{(\tau_2)}$  which will be used later in this work. Thus, from the multinomial distributions indicated in Subsection 3.1 it follows that  $R_x^{(\tau_1)}|S_A^{(\tau_1)} \sim \text{Bin}(\tau_1 - M^{(\tau_1)}, \pi_x^{(1)})$ ,  $R_x^{(A_i)}|M_i^{(\tau_1)} \sim \text{Bin}(M_i^{(\tau_1)}, \pi_x^{(A_i)})$  and  $R_x^{(\tau_2)}|S_A^{(\tau_1)} \sim \text{Bin}(\tau_2, \pi_x^{(2)})$ ; therefore  $R_x^{(\tau_1)}|\tau_1 \sim \text{Bin}\left[\tau_1, (1 - n/N) \pi_x^{(1)}\right]$ ,  $R_x^{(A_i)}|\tau_1 \sim \text{Bin}\left(\tau_1, \pi_x^{(A_i)}/N\right)$  and  $R_x^{(\tau_2)}|\tau_2 \sim \text{Bin}(\tau_2, \pi_x^{(2)})$ .

## 4.2 Asymptotic multivariate normal distribution of estimators of $(\tau_1, \boldsymbol{\theta}_1^*)$

**Theorem 1.** Let  $\boldsymbol{\theta}_1^* = (\theta_1^*, \dots, \theta_{q_1}^*)$  be the true value of  $\boldsymbol{\theta}_1$ . Let  $\hat{\tau}_1^{(\tau_1)}$  and  $\hat{\boldsymbol{\theta}}_1^{(\tau_1)} = (\hat{\theta}_{11}^{(\tau_1)}, \dots, \hat{\theta}_{1q_1}^{(\tau_1)})$  be estimators of  $\tau_1$  and  $\boldsymbol{\theta}_1^*$ , such that

- (i)  $\hat{\boldsymbol{\theta}}_1^{(\tau_1)} \xrightarrow{P} \boldsymbol{\theta}_1^*$ .
- (ii)  $\tau_1^{-1/2} \left\{ \hat{\tau}_1^{(\tau_1)} - \left( M^{(\tau_1)} + R_1^{(\tau_1)} \right) / \left[ 1 - (1 - n/N)\pi_0^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) \right] \right\} \xrightarrow{P} 0$ .
- (iii)  $\tau_1^{-1/2} \left[ \frac{\partial}{\partial \theta_j^{(1)}} l_{(1)}^{(\tau_1)}(\hat{\tau}_1^{(\tau_1)}, \hat{\boldsymbol{\theta}}_1^{(\tau_1)}) \right] \xrightarrow{P} 0, j = 1, \dots, q_1$ .

In addition, let  $\boldsymbol{\Sigma}_1^{-1}$  be the  $(q_1 + 1) \times (q_1 + 1)$  matrix whose elements are

$$\begin{aligned} [\boldsymbol{\Sigma}_1^{-1}]_{1,1} &= \left[ 1 - (1 - n/N)\pi_0^{(1)}(\boldsymbol{\theta}_1^*) \right] / \left[ (1 - n/N)\pi_0^{(1)}(\boldsymbol{\theta}_1^*) \right], \\ [\boldsymbol{\Sigma}_1^{-1}]_{1,j+1} &= [\boldsymbol{\Sigma}_1^{-1}]_{j+1,1} = - \left[ 1/\pi_0^{(1)}(\boldsymbol{\theta}_1^*) \right] \left[ \partial \pi_0^{(1)}(\boldsymbol{\theta}_1^*) / \partial \theta_j^{(1)} \right], \quad j = 1, \dots, q_1, \\ [\boldsymbol{\Sigma}_1^{-1}]_{i+1,j+1} &= [\boldsymbol{\Sigma}_1^{-1}]_{j+1,i+1} = \left( 1 - \frac{n}{N} \right) \sum_{\mathbf{x} \in \Omega} \left[ 1/\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*) \right] \left[ \partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*) / \partial \theta_i^{(1)} \right] \left[ \partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*) / \partial \theta_j^{(1)} \right] \\ &\quad + \frac{1}{N} \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \left[ 1/\pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*) \right] \left[ \partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*) / \partial \theta_i^{(1)} \right] \left[ \partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*) / \partial \theta_j^{(1)} \right], \\ &\quad i, j = 1, \dots, q_1, \end{aligned}$$

and which is assumed to be a non-singular matrix. Then

$$\left[ \tau_1^{-1/2} \left( \hat{\tau}_1^{(\tau_1)} - \tau_1 \right), \tau_1^{1/2} \left( \hat{\boldsymbol{\theta}}_1^{(\tau_1)} - \boldsymbol{\theta}_1^* \right) \right] \xrightarrow{D} N_{q_1+1}(\mathbf{0}, \boldsymbol{\Sigma}_1),$$

where  $\boldsymbol{\Sigma}_1$  is the inverse of  $\boldsymbol{\Sigma}_1^{-1}$  and  $\mathbf{0} = (0, \dots, 0) \in \mathbb{R}^{q_1+1}$ .

*Proof.* Evaluating equation (4) at  $(\hat{\tau}_1^{(\tau_1)}, \hat{\boldsymbol{\theta}}_1^{(\tau_1)})$  we get

$$\begin{aligned}
& \frac{\partial}{\partial \theta_j^{(1)}} l_{(1)}^{(\tau_1)} \left( \hat{\tau}_1^{(\tau_1)}, \hat{\boldsymbol{\theta}}_1^{(\tau_1)} \right) \\
&= \sum_{\mathbf{x} \in \Omega - \mathbf{0}} \frac{R_{\mathbf{x}}^{(\tau_1)}}{\pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} + \frac{\hat{\tau}_1^{(\tau_1)} - M^{(\tau_1)} - R_1^{(\tau_1)}}{\pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \\
&+ \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \frac{R_{\mathbf{x}}^{(A_l, \tau_1)}}{\pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \\
&= \sum_{\mathbf{x} \in \Omega} \frac{R_{\mathbf{x}}^{(\tau_1)}}{\pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} + \frac{\left[ \hat{\tau}_1^{(\tau_1)} - M^{(\tau_1)} - R_1^{(\tau_1)} \right] - \left[ \tau_1 - M^{(\tau_1)} - R_1^{(\tau_1)} \right]}{\pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \\
&+ \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \frac{R_{\mathbf{x}}^{(A_l, \tau_1)}}{\pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \\
&= \sum_{\mathbf{x} \in \Omega} \frac{R_{\mathbf{x}}^{(\tau_1)}}{\pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} + \frac{\hat{\tau}_1^{(\tau_1)} - \tau_1}{\pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} + \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \frac{R_{\mathbf{x}}^{(A_l, \tau_1)}}{\pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}}.
\end{aligned} \tag{10}$$

Since

$$\sum_{\mathbf{x} \in \Omega} \partial \pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) / \partial \theta_j^{(1)} = 0 \quad \text{and} \quad \sum_{\mathbf{x} \in \Omega_{-l}} \partial \pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) / \partial \theta_j^{(1)} = 0, \tag{11}$$

from (10) we get that

$$\begin{aligned}
& \tau_1^{-1/2} \left\{ \sum_{\mathbf{x} \in \Omega} \frac{R_{\mathbf{x}}^{(\tau_1)} - (\tau_1 - M^{(\tau_1)}) \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \right. \\
&+ \left. \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \frac{R_{\mathbf{x}}^{(A_l, \tau_1)} - M_l^{(\tau_1)} \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)}{\pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \right\} - \tau_1^{-1/2} \frac{\partial}{\partial \theta_j^{(1)}} l_{(1)}^{(\tau_1)} \left( \hat{\tau}_1^{(\tau_1)}, \hat{\boldsymbol{\theta}}_1^{(\tau_1)} \right) \\
&= -\tau_1^{-1/2} (\hat{\tau}_1^{(\tau_1)} - \tau_1) \left[ \frac{1}{\pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \right] \\
&+ \tau_1^{1/2} \left\{ \frac{\tau_1 - M^{(\tau_1)}}{\tau_1} \sum_{\mathbf{x} \in \Omega} \frac{\pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) - \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \right. \\
&+ \left. \sum_{l=1}^n \frac{M_l^{(\tau_1)}}{\tau_1} \sum_{\mathbf{x} \in \Omega_{-l}} \frac{\pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) - \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)}{\pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \right\}.
\end{aligned} \tag{12}$$

Let  $Y_{\mathbf{x}}^{(\tau_1)} = R_{\mathbf{x}}^{(\tau_1)} - (\tau_1 - M^{(\tau_1)})\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)$ ,  $Y_{\mathbf{x}}^{(A_l, \tau_1)} = R_{\mathbf{x}}^{(A_l, \tau_1)} - M_l^{(\tau_1)}\pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)$  and

$$\begin{aligned} Z_{j+1}^{(\tau_1)} &= \tau_1^{-1/2} \left[ \sum_{\mathbf{x} \in \Omega} \frac{R_{\mathbf{x}}^{(\tau_1)}}{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} + \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \frac{R_{\mathbf{x}}^{(A_l, \tau_1)}}{\pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \right] \\ &= \tau_1^{-1/2} \left[ \sum_{\mathbf{x} \in \Omega} \frac{Y_{\mathbf{x}}^{(\tau_1)}}{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} + \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \frac{Y_{\mathbf{x}}^{(A_l, \tau_1)}}{\pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \right], \end{aligned}$$

where the last equality is obtained using (11) but replacing  $\hat{\boldsymbol{\theta}}_1^{(\tau_1)}$  by  $\boldsymbol{\theta}_1^*$ . Then, the difference between the left-hand side of (12) and  $Z_{j+1}^{(\tau_1)}$  is given by

$$\begin{aligned} &\tau_1^{-1/2} \left\{ \sum_{\mathbf{x} \in \Omega} \frac{Y_{\mathbf{x}}^{(\tau_1)}}{\pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} + \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \frac{Y_{\mathbf{x}}^{(A_l, \tau_1)}}{\pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \right\} \\ &- \tau_1^{-1/2} \frac{\partial}{\partial \theta_j^{(1)}} l_{(1)}^{(\tau_1)} \left( \hat{\tau}_1^{(\tau_1)}, \hat{\boldsymbol{\theta}}_1^{(\tau_1)} \right) - Z_{j+1}^{(\tau_1)} \\ &= \tau_1^{-1/2} \left\{ \sum_{\mathbf{x} \in \Omega} Y_{\mathbf{x}}^{(\tau_1)} \left[ \frac{1}{\pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} - \frac{1}{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \right] \right. \\ &\quad \left. + \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} Y_{\mathbf{x}}^{(A_l, \tau_1)} \left[ \frac{1}{\pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} - \frac{1}{\pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \right] \right\} \\ &- \tau_1^{-1/2} \frac{\partial}{\partial \theta_j^{(1)}} l_{(1)}^{(\tau_1)} \left( \hat{\tau}_1^{(\tau_1)}, \hat{\boldsymbol{\theta}}_1^{(\tau_1)} \right). \end{aligned} \quad (13)$$

Since unconditionally  $E(Y_{\mathbf{x}}^{(\tau_1)}) = 0$  and  $V(Y_{\mathbf{x}}^{(\tau_1)}) = \tau_1(1 - n/N)\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)[1 - \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)]$ , and also  $E(Y_{\mathbf{x}}^{(A_l, \tau_1)}) = 0$  and  $V(Y_{\mathbf{x}}^{(A_l, \tau_1)}) = \tau_1(1/N)\pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)[1 - \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)]$ , it follows that  $\tau_1^{-1/2}Y_{\mathbf{x}}^{(\tau_1)} = O_p(1)$  and  $\tau_1^{-1/2}Y_{\mathbf{x}}^{(A_l, \tau_1)} = O_p(1)$ . Consequently, these results along with conditions (3)-(4) and conditions (i) and (iii) of the theorem imply that (13) converges to zero in probability.

On the other hand, by the mean value theorem for functions of several variables we have that

$$\begin{aligned} \pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) - \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*) &= \sum_{i=1}^{q_1} \left( \hat{\theta}_{1i}^{(\tau_1)} - \theta_{1i}^* \right) \partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_{1\mathbf{x}}^{(\tau_1)}) / \partial \theta_i^{(1)} \quad \text{and} \quad (14) \\ \pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) - \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*) &= \sum_{i=1}^{q_1} \left( \hat{\theta}_{1i}^{(\tau_1)} - \theta_{1i}^* \right) \partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_{A_l\mathbf{x}}^{(\tau_1)}) / \partial \theta_i^{(1)}, \end{aligned}$$

where  $\boldsymbol{\theta}_{1\mathbf{x}}^{(\tau_1)}$  and  $\boldsymbol{\theta}_{A_l\mathbf{x}}^{(\tau_1)}$  are between  $\hat{\boldsymbol{\theta}}_1^{(\tau_1)}$  and  $\boldsymbol{\theta}_1^*$ . Since the difference between the right-hand

side of (12) and  $Z_{j+1}^{(\tau_1)}$  also converges to zero in probability, we have that

$$\begin{aligned}
& -\tau_1^{-1/2}(\hat{\tau}_1^{(\tau_1)} - \tau_1) \left[ \frac{1}{\pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \right] \\
& + \tau_1^{1/2} \left\{ \frac{\tau_1 - M^{(\tau_1)}}{\tau_1} \sum_{\mathbf{x} \in \Omega} \frac{1}{\pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \sum_{i=1}^{q_1} (\hat{\theta}_{1i}^{(\tau_1)} - \theta_{1i}^*) \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_{1\mathbf{x}}^{(\tau_1)})}{\partial \theta_i^{(1)}} \right. \\
& \left. + \sum_{l=1}^n \frac{M_l^{(\tau_1)}}{\tau_1} \sum_{\mathbf{x} \in \Omega_{-l}} \frac{1}{\pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \sum_{i=1}^{q_1} (\hat{\theta}_{1i}^{(\tau_1)} - \theta_{1i}^*) \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_{A_l\mathbf{x}}^{(\tau_1)})}{\partial \theta_i^{(1)}} \right\} - Z_{j+1}^{(\tau_1)} \\
& = \left[ \hat{\Sigma}_1^{-1} \right]_{j+1,1} \left[ \tau_1^{-1/2}(\hat{\tau}_1^{(\tau_1)} - \tau_1) \right] + \sum_{i=1}^{q_1} \left[ \hat{\Sigma}_1^{-1} \right]_{j+1,i+1} \left[ \tau_1^{1/2} (\hat{\theta}_{1i}^{(\tau_1)} - \theta_{1i}^*) \right] - Z_{j+1}^{(\tau_1)} \xrightarrow{P} 0,
\end{aligned} \tag{15}$$

where

$$\begin{aligned}
\left[ \hat{\Sigma}_1^{-1} \right]_{j+1,1} &= -\frac{1}{\pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \quad \text{and} \\
\left[ \hat{\Sigma}_1^{-1} \right]_{j+1,i+1} &= \frac{\tau_1 - M^{(\tau_1)}}{\tau_1} \sum_{\mathbf{x} \in \Omega} \frac{1}{\pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_{1\mathbf{x}}^{(\tau_1)})}{\partial \theta_i^{(1)}} \\
&+ \sum_{l=1}^n \frac{M_l^{(\tau_1)}}{\tau_1} \sum_{\mathbf{x} \in \Omega_{-l}} \frac{1}{\pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_{A_l\mathbf{x}}^{(\tau_1)})}{\partial \theta_i^{(1)}}. \tag{16}
\end{aligned}$$

Expression (5) suggests the following equality in terms of  $\hat{\tau}_1^{(\tau_1)} - \tau_1$  and  $\pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)$ :

$$\begin{aligned}
& \tau_1^{-1/2} \left\{ \hat{\tau}_1^{(\tau_1)} \left[ 1 - (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) \right] - \left( M^{(\tau_1)} + R_1^{(\tau_1)} \right) \right\} = \tau_1^{-1/2} \left( \hat{\tau}_1^{(\tau_1)} - \tau_1 \right) \\
& \times \left[ 1 - (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) \right] - \tau_1^{-1/2} \left\{ \left( M^{(\tau_1)} + R_1^{(\tau_1)} \right) - \tau_1 \left[ 1 - (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right] \right\} \\
& - \tau_1^{1/2} (1 - n/N) \left[ \pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right].
\end{aligned}$$

By condition (ii) of the theorem it follows that the left hand-side of the previous equation converges to zero in probability. Therefore, if we divide the right hand-side of this equation by  $(1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)$  and use (14), we will get that the following expression also converges to zero in probability, that is



$$\begin{aligned}
& \tau_1^{-1/2} \left( \hat{\tau}_1^{(\tau_1)} - \tau_1 \right) \frac{1 - (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{(1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)} \\
& - \tau_1^{-1/2} \frac{\left( M^{(\tau_1)} + R_1^{(\tau_1)} \right) - \tau_1 \left[ 1 - (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right]}{(1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)} \\
& - \sum_{i=1}^{q_1} \tau_1^{1/2} \left( \hat{\theta}_{1i}^{(\tau_1)} - \theta_{1i}^* \right) \frac{1}{\pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_{10}^{(\tau_1)})}{\partial \theta_i^{(1)}}, \\
& = \left[ \hat{\boldsymbol{\Sigma}}_1^{-1} \right]_{1,1} \left[ \tau_1^{-1/2} \left( \hat{\tau}_1^{(\tau_1)} - \tau_1 \right) \right] + \sum_{i=1}^{q_1} \left[ \hat{\boldsymbol{\Sigma}}_1^{-1} \right]_{1,i+1} \left[ \tau_1^{1/2} \left( \hat{\theta}_{1i}^{(\tau_1)} - \theta_{1i}^* \right) \right] - Z_1^{(\tau_1)} \xrightarrow{P} 0, \quad (17)
\end{aligned}$$

where  $\boldsymbol{\theta}_{10}^{(\tau_1)}$  is between  $\hat{\boldsymbol{\theta}}_1^{(\tau_1)}$  and  $\boldsymbol{\theta}_1^*$  and

$$\left[ \hat{\boldsymbol{\Sigma}}_1^{-1} \right]_{1,1} = \frac{1 - (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{(1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}, \quad \left[ \hat{\boldsymbol{\Sigma}}_1^{-1} \right]_{1,i+1} = -\frac{1}{\pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_{10}^{(\tau_1)})}{\partial \theta_i^{(1)}} \quad (18)$$

and

$$Z_1^{(\tau_1)} = \tau_1^{-1/2} \frac{\left( M^{(\tau_1)} + R_1^{(\tau_1)} \right) - \tau_1 \left[ 1 - (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right]}{(1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}.$$

Let  $\mathbf{W}_1^{(\tau_1)} = [\tau_1^{-1/2}(\hat{\tau}_1^{(\tau_1)} - \tau_1), \tau_1^{1/2}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)} - \boldsymbol{\theta}_1^*)]'$  and  $\mathbf{Z}^{(\tau_1)} = [Z_1^{(\tau_1)}, Z_2^{(\tau_1)}, \dots, Z_{q_1+1}^{(\tau_1)}]'$ , by the previous results we have that

$$\hat{\boldsymbol{\Sigma}}_1^{-1} \mathbf{W}_1^{(\tau_1)} - \mathbf{Z}^{(\tau_1)} \xrightarrow{P} \mathbf{0}, \quad (19)$$

where  $\hat{\boldsymbol{\Sigma}}_1^{-1}$  is the  $(q_1 + 1) \times (q_1 + 1)$  matrix whose elements are defined in (16) and (18). Notice that from the definitions of the matrices  $\boldsymbol{\Sigma}_1^{-1}$  and  $\hat{\boldsymbol{\Sigma}}_1^{-1}$ , conditions **(3)**-**(4)** and condition **(i)** of the theorem along with the fact that  $(\tau_1 - M^{(\tau_1)})/\tau_1 \xrightarrow{P} 1 - n/N$  and  $M_t^{(\tau_1)}/\tau_1 \xrightarrow{P} 1/N$ , it follows that  $\hat{\boldsymbol{\Sigma}}_1^{-1} \xrightarrow{P} \boldsymbol{\Sigma}_1^{-1}$ .

We will show that  $\mathbf{Z}^{(\tau_1)} \xrightarrow{D} \mathbf{Z} \sim N_{q_1+1}(\mathbf{0}, \boldsymbol{\Sigma}_1^{-1})$  as  $\tau_1 \rightarrow \infty$ . To do this, we will associate with each element  $t \in U_1$ ,  $t = 1, \dots, \tau_1$ , a random vector  $\mathbf{V}_t^{(1)} = [V_{t,1}^{(1)}, \dots, V_{t,q_1+1}^{(1)}]'$  such that

- (a)  $V_{t,1}^{(1)} = 1$  and  $V_{t,j+1}^{(1)} = [\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)]^{-1} \partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*) / \partial \theta_j^{(1)}$ ,  $j = 1, \dots, q_1$ , if  $t \in U_1 - S_0$  and its associated vector  $X_t^{(1)}$  of link-indicator variables equals the vector  $\mathbf{x} \in \Omega - \{\mathbf{0}\}$ ;
- (b)  $V_{t,1}^{(1)} = - \left[ 1 - (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right] / \left[ (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right]$  and  $V_{t,j+1}^{(1)} = [\pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)]^{-1} \times \partial \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) / \partial \theta_j^{(1)}$ ,  $j = 1, \dots, q_1$ , if  $t \in U_1 - S_0$  and its associated vector  $X_t^{(1)}$  of link-indicator variables equals the vector  $\mathbf{0} \in \Omega$ , and

- (c)  $V_{t,1}^{(1)} = 1$  and  $V_{t,j+1}^{(1)} = [\pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)]^{-1} \partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*) / \partial \theta_j^{(1)}$ ,  $j = 1, \dots, q_1$ , if  $t \in A_l \in S_A$  and its associated vector  $X_t^{(1)}$  of link-indicator variables equals the vector  $\mathbf{x} \in \Omega_{-l}$ .

Since

$$\begin{aligned} \tau_1^{-1/2} \sum_{t=1}^{\tau_1} V_{t,1}^{(1)} &= \tau_1^{-1/2} \left[ \left( M^{(\tau_1)} + R_1^{(\tau_1)} \right) - \left( \tau_1 - M^{(\tau_1)} - R_1^{(\tau_1)} \right) \frac{1 - (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{(1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)} \right] \\ &= Z_1^{(\tau_1)}, \end{aligned}$$

and

$$\begin{aligned} \tau_1^{-1/2} \sum_{t=1}^{\tau_1} V_{t,j+1}^{(1)} &= \tau_1^{-1/2} \left[ \sum_{\mathbf{x} \in \Omega} \frac{R_{\mathbf{x}}^{(\tau_1)}}{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} + \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \frac{R_{\mathbf{x}}^{(A_l, \tau_1)}}{\pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \right] = Z_{j+1}^{(\tau_1)} \\ &\quad j = 1, \dots, q_1; \end{aligned}$$

it follows that  $\mathbf{Z}^{(\tau_1)} = \tau_1^{-1/2} \sum_{t=1}^{\tau_1} \mathbf{V}_t^{(1)}$ .

From the definition of  $V_{t,j}^{(1)}$  we have that

$$\begin{aligned} \Pr \left\{ V_{t,1}^{(1)} = 1 \right\} &= (1 - n/N) \left[ 1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right] + n/N, \\ \Pr \left\{ V_{t,1}^{(1)} = - \left[ 1 - (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right] / \left[ (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right] \right\} &= (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*), \\ \Pr \left\{ V_{t,j+1}^{(1)} = [\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)]^{-1} \partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*) / \partial \theta_j^{(1)} \right\} &= (1 - n/N) \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*), \mathbf{x} \in \Omega, j = 1, \dots, q_1, \text{ and} \\ \Pr \left\{ V_{t,j+1}^{(1)} = [\pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)]^{-1} \partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*) / \partial \theta_j^{(1)} \right\} &= (1/N) \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*), \mathbf{x} \in \Omega_{-l}, j = 1, \dots, q_1, \\ &\quad l = 1, \dots, n; \end{aligned}$$

therefore, the expected values of the variables  $V_{t,j}^{(1)}$  are

$$E \left( V_{t,1}^{(1)} \right) = (1 - n/N) \left[ 1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right] + n/N - \left[ 1 - (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right] = 0$$

and

$$\begin{aligned} E \left( V_{t,j+1}^{(1)} \right) &= \sum_{\mathbf{x} \in \Omega} \partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*) / \partial \theta_j^{(1)} (1 - n/N) + \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*) / \partial \theta_j^{(1)} (1/N) = 0, \\ &\quad j = 1, \dots, q_1, \end{aligned}$$

because of (11). Thus,  $E \left( \mathbf{V}_t^{(1)} \right) = \mathbf{0}$ ,  $t = 1, \dots, \tau_1$ . Furthermore, their variances are

$$\begin{aligned} V \left( V_{t,1}^{(1)} \right) &= (1 - n/N) \left[ 1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right] + n/N + \frac{\left[ 1 - (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right]^2}{(1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)} \\ &= \frac{1 - (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{(1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)} \end{aligned}$$

and

$$\begin{aligned} V(V_{t,j+1}^{(1)}) &= (1 - n/N) \sum_{\mathbf{x} \in \Omega} \frac{1}{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)} \left[ \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \right]^2 \\ &\quad + \frac{1}{N} \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \frac{1}{\pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)} \left[ \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \right]^2, \quad j = 1, \dots, q_1, \end{aligned}$$

and their covariances are

$$\begin{aligned} Cov(V_{t,1}^{(1)}, V_{t,j+1}^{(1)}) &= \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} (1 - n/N) - \frac{1 - (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{(1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \\ &\quad \times (1 - n/N) + \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \frac{1}{N} \\ &= -\frac{1}{\pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}}, \quad j = 1, \dots, q_1, \quad \text{and} \\ Cov(V_{t,j+1}^{(1)}, V_{t,j'+1}^{(1)}) &= (1 - n/N) \sum_{\mathbf{x} \in \Omega} \frac{1}{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_{j'}^{(1)}} \\ &\quad + \frac{1}{N} \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \frac{1}{\pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)}{\partial \theta_{j'}^{(1)}}, \\ &\quad j, j' = 1, \dots, q_1, \quad j \neq j'. \end{aligned}$$

Therefore, the variance-covariance matrix of  $\mathbf{V}_t^{(1)}$  is  $\Sigma_1^{-1}$ .

Finally, since the  $\mathbf{V}_t^{(1)}$ ,  $t = 1, \dots, \tau_1$ , are independent and identically distributed random vectors, by the central limit theorem it follows that

$$\mathbf{Z}^{(\tau_1)} = \tau_1^{-1/2} \sum_{t=1}^{\tau_1} \mathbf{V}_t^{(1)} \xrightarrow{D} \mathbf{Z} \sim N_{q_1+1}(\mathbf{0}, \Sigma_1^{-1}).$$

Consequently by (19),

$$\mathbf{W}_1^{(\tau_1)} = \left[ \tau_1^{-1/2} \left( \hat{\tau}_1^{(\tau_1)} - \tau_1 \right), \tau_1^{1/2} \left( \hat{\boldsymbol{\theta}}_1^{(\tau_1)} - \boldsymbol{\theta}_1^* \right) \right] \xrightarrow{D} \Sigma_1 \mathbf{Z} \sim N_{q_1+1}(\mathbf{0}, \Sigma_1)$$

as  $\hat{\Sigma}_1 \xrightarrow{P} \Sigma_1$ . □

### 4.3 Asymptotic multivariate normal distribution of estimators of $\boldsymbol{\theta}_1^*$

**Theorem 2.** Let  $\boldsymbol{\theta}_1^* = (\theta_1^*, \dots, \theta_{q_1}^*)$  be the true value of  $\boldsymbol{\theta}_1$ . Let  $\hat{\boldsymbol{\theta}}_1^{(\tau_1)} = (\hat{\theta}_{11}^{(\tau_1)}, \dots, \hat{\theta}_{1q_1}^{(\tau_1)})$  be an estimator of  $\boldsymbol{\theta}_1^*$ , such that

(i)  $\hat{\boldsymbol{\theta}}_1^{(\tau_1)} \xrightarrow{P} \boldsymbol{\theta}_1^*$ .

$$(ii) \quad \tau_1^{-1/2} \left\{ \frac{\partial}{\partial \theta_j^{(1)}} \ln \left[ L_{11}^{(\tau_1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) L_0^{(\tau_1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) \right] \right\} \xrightarrow{P} 0, j = 1, \dots, q_1.$$

In addition, let  $\boldsymbol{\Psi}_1^{-1}$  be the  $q_1 \times q_1$  matrix whose elements are

$$\begin{aligned} [\boldsymbol{\Psi}_1^{-1}]_{i,j} &= [\boldsymbol{\Psi}_1^{-1}]_{j,i} = (1 - n/N)[1 - \pi_0^{(1)}(\boldsymbol{\theta}_1^*)] \\ &\quad \times \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} [1/\tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)] \left[ \partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*) / \partial \theta_i^{(1)} \right] \left[ \partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*) / \partial \theta_j^{(1)} \right] \\ &\quad + \frac{1}{N} \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} [1/\pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)] \left[ \partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*) / \partial \theta_i^{(1)} \right] \left[ \partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*) / \partial \theta_j^{(1)} \right], \\ &\quad i, j = 1, \dots, q_1, \end{aligned}$$

where  $\tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*) = \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)/[1 - \pi_0^{(1)}(\boldsymbol{\theta}_1^*)]$ ,  $\mathbf{x} \in \Omega - \{\mathbf{0}\}$ , and suppose that  $\boldsymbol{\Psi}_1^{-1}$  is a non-singular matrix. Then

$$\tau_1^{1/2} \left[ \hat{\boldsymbol{\theta}}_1^{(\tau_1)} - \boldsymbol{\theta}_1^* \right] \xrightarrow{D} N_{q_1}(\mathbf{0}, \boldsymbol{\Psi}_1),$$

where  $\boldsymbol{\Psi}_1$  is the inverse of  $\boldsymbol{\Psi}_1^{-1}$  and  $\mathbf{0} = (0, \dots, 0) \in \mathbb{R}^{q_1}$ .

Furthermore, if  $\hat{\tau}_1^{(\tau_1)}$  is an estimator of  $\tau_1$  such that

$$(iii) \quad \tau_1^{-1/2} \left\{ \hat{\tau}_1^{(\tau_1)} - \left( M^{(\tau_1)} + R_1^{(\tau_1)} \right) / \left[ 1 - (1 - n/N)\pi_0^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) \right] \right\} \xrightarrow{P} 0,$$

then

$$\tau_1^{-1/2} \left( \hat{\tau}_1^{(\tau_1)} - \tau_1 \right) \xrightarrow{D} N(0, \sigma_1^2),$$

where

$$\sigma_1^2 = \frac{1 - n/N}{1 - (1 - n/N)\pi_0^{(1)}(\boldsymbol{\theta}_1^*)} \left\{ \pi_0^{(1)}(\boldsymbol{\theta}_1^*) + \frac{(1 - n/N) \left[ \nabla \pi_0^{(1)}(\boldsymbol{\theta}_1^*) \right]' \boldsymbol{\Psi}_1 \left[ \nabla \pi_0^{(1)}(\boldsymbol{\theta}_1^*) \right]}{1 - (1 - n/N)\pi_0^{(1)}(\boldsymbol{\theta}_1^*)} \right\}, \quad (20)$$

and  $\nabla \pi_0^{(1)}(\boldsymbol{\theta}_1^*) = \left[ \partial \pi_0^{(1)}(\boldsymbol{\theta}_1^*) / \partial \theta_1^{(1)}, \dots, \partial \pi_0^{(1)}(\boldsymbol{\theta}_1^*) / \partial \theta_{q_1}^{(1)} \right]'$  is the gradient of  $\pi_0^{(1)}(\boldsymbol{\theta}_1)$  evaluated at  $\boldsymbol{\theta}_1^*$ .

*Proof.* From the definitions of  $L_{11}^{(\tau_1)}(\boldsymbol{\theta}_1)$  and  $L_0^{(\tau_1)}(\boldsymbol{\theta}_1)$  we have that

$$\begin{aligned} \frac{\partial}{\partial \theta_j^{(1)}} \ln \left[ L_{11}^{(\tau_1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) L_0^{(\tau_1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) \right] &= \sum_{\mathbf{x} \in \Omega - \mathbf{0}} \frac{R_{\mathbf{x}}^{(\tau_1)}}{\tilde{\pi}_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \\ &\quad + \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \frac{R_{\mathbf{x}}^{(A_l, \tau_1)}}{\pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}}. \end{aligned} \quad (21)$$

Since

$$\sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) / \partial \theta_j^{(1)} = 0 \quad \text{and} \quad \sum_{\mathbf{x} \in \Omega_{-l}} \partial \pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) / \partial \theta_j^{(1)} = 0, \quad (22)$$

from (21) we get that

$$\begin{aligned}
& \tau_1^{-1/2} \left\{ \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{R_{\mathbf{x}}^{(\tau_1)} - R_1^{(\tau_1)} \tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\tilde{\pi}_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \right. \\
& + \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \frac{R_{\mathbf{x}}^{(A_l, \tau_1)} - M_l^{(\tau_1)} \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)}{\pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \left. \right\} \\
& - \frac{\partial}{\partial \theta_j^{(1)}} \ln \left[ L_{11}^{(\tau_1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) L_0^{(\tau_1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) \right] \\
& = \tau_1^{1/2} \left\{ \frac{R_1^{(\tau_1)}}{\tau_1} \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{\tilde{\pi}_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) - \tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\tilde{\pi}_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \right. \\
& + \sum_{l=1}^n \frac{M_l^{(\tau_1)}}{\tau_1} \sum_{\mathbf{x} \in \Omega_{-l}} \frac{\pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) - \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)}{\pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \left. \right\}. \tag{23}
\end{aligned}$$

Let  $Y_{\mathbf{x}}^{(\tau_1)} = R_{\mathbf{x}}^{(\tau_1)} - R_1^{(\tau_1)} \tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)$ ,  $Y_{\mathbf{x}}^{(A_l, \tau_1)} = R_{\mathbf{x}}^{(A_l, \tau_1)} - M_l^{(\tau_1)} \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)$  and

$$\begin{aligned}
Z_j^{(\tau_1)} &= \tau_1^{-1/2} \left[ \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{R_{\mathbf{x}}^{(\tau_1)}}{\tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} + \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \frac{R_{\mathbf{x}}^{(A_l, \tau_1)}}{\pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \right] \\
&= \tau_1^{-1/2} \left[ \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{Y_{\mathbf{x}}^{(\tau_1)}}{\tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} + \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \frac{Y_{\mathbf{x}}^{(A_l, \tau_1)}}{\pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \right],
\end{aligned}$$

where the last equality is obtained using (22) but replacing  $\hat{\boldsymbol{\theta}}_1^{(\tau_1)}$  by  $\boldsymbol{\theta}_1^*$ . Then, the difference between the left-hand side of (23) and  $Z_j^{(\tau_1)}$  is given by

$$\begin{aligned}
& \tau_1^{-1/2} \left\{ \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{Y_{\mathbf{x}}^{(\tau_1)}}{\tilde{\pi}_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} + \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \frac{Y_{\mathbf{x}}^{(A_l, \tau_1)}}{\pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \right\} \\
& - \tau_1^{-1/2} \frac{\partial}{\partial \theta_j^{(1)}} \ln \left[ L_{11}^{(\tau_1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) L_0^{(\tau_1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) \right] - Z_j^{(\tau_1)} \\
& = \tau_1^{-1/2} \left\{ \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} Y_{\mathbf{x}}^{(\tau_1)} \left[ \frac{1}{\tilde{\pi}_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} - \frac{1}{\tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \right] \right. \\
& + \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} Y_{\mathbf{x}}^{(A_l, \tau_1)} \left[ \frac{1}{\pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} - \frac{1}{\pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \right] \left. \right\} \\
& - \tau_1^{-1/2} \frac{\partial}{\partial \theta_j^{(1)}} \ln \left[ L_{11}^{(\tau_1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) L_0^{(\tau_1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) \right]. \tag{24}
\end{aligned}$$

Since  $\tau_1^{-1/2} Y_{\mathbf{x}}^{(\tau_1)} = O_p(1)$  and  $\tau_1^{-1/2} Y_{\mathbf{x}}^{(A_l, \tau_1)} = O_p(1)$ , these results along with conditions (3)-(4) and conditions (i) and (ii) of the theorem imply that (24) converges to zero in probability.

On the other hand, by the mean value theorem of several variables we have that

$$\begin{aligned}\tilde{\pi}_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) - \tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*) &= \sum_{i=1}^{q_1} \left( \hat{\theta}_{1i}^{(\tau_1)} - \theta_{1i}^* \right) \partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_{1\mathbf{x}}^{(\tau_1)}) / \partial \theta_i^{(1)} \quad \text{and} \quad (25) \\ \pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)}) - \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*) &= \sum_{i=1}^{q_1} \left( \hat{\theta}_{1i}^{(\tau_1)} - \theta_{1i}^* \right) \partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_{A_l\mathbf{x}}^{(\tau_1)}) / \partial \theta_i^{(1)},\end{aligned}$$

where  $\boldsymbol{\theta}_{1\mathbf{x}}^{(\tau_1)}$  and  $\boldsymbol{\theta}_{A_l\mathbf{x}}^{(\tau_1)}$  are between  $\hat{\boldsymbol{\theta}}_1^{(\tau_1)}$  and  $\boldsymbol{\theta}_1^*$ . Since the difference between the right-hand side of (23) and  $Z_j^{(\tau_1)}$  also converges to zero in probability, we have that

$$\begin{aligned}& \tau_1^{1/2} \left\{ \frac{R_1^{(\tau_1)}}{\tau_1} \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{1}{\tilde{\pi}_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \sum_{i=1}^{q_1} \left( \hat{\theta}_{1i}^{(\tau_1)} - \theta_{1i}^* \right) \frac{\partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_{1\mathbf{x}}^{(\tau_1)})}{\partial \theta_i^{(1)}} \right. \\ & \left. + \sum_{l=1}^n \frac{M_l^{(\tau_1)}}{\tau_1} \sum_{\mathbf{x} \in \Omega_{-l}} \frac{1}{\pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \sum_{i=1}^{q_1} \left( \hat{\theta}_{1i}^{(\tau_1)} - \theta_{1i}^* \right) \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_{A_l\mathbf{x}}^{(\tau_1)})}{\partial \theta_i^{(1)}} \right\} - Z_j^{(\tau_1)} \\ &= \sum_{i=1}^{q_1} [\hat{\Psi}_1^{-1}]_{j,i} \left[ \tau_1^{1/2} \left( \hat{\theta}_{1i}^{(\tau_1)} - \theta_{1i}^* \right) \right] - Z_j^{(\tau_1)} \xrightarrow{P} 0, \quad (26)\end{aligned}$$

where

$$\begin{aligned}[\hat{\Psi}_1^{-1}]_{j,i} &= \frac{R_1^{(\tau_1)}}{\tau_1} \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{1}{\tilde{\pi}_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \frac{\partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_{1\mathbf{x}}^{(\tau_1)})}{\partial \theta_i^{(1)}} \\ &+ \sum_{l=1}^n \frac{M_l^{(\tau_1)}}{\tau_1} \sum_{\mathbf{x} \in \Omega_{-l}} \frac{1}{\pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{\partial \theta_j^{(1)}} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_{A_l\mathbf{x}}^{(\tau_1)})}{\partial \theta_i^{(1)}}. \quad (27)\end{aligned}$$

Notice that from the definitions of the matrices  $\Psi_1^{-1}$  and  $\hat{\Psi}_1^{-1}$ , conditions (3)-(4) and condition (i) of the theorem along with the fact that  $R_1^{(\tau_1)}/\tau_1 \xrightarrow{P} (1 - n/N)[1 - \pi_0^{(1)}(\boldsymbol{\theta}_1^*)]$  and  $M_l^{(\tau_1)}/\tau_1 \xrightarrow{P} 1/N$ , it follows that  $\hat{\Psi}_1^{-1} \xrightarrow{P} \Psi_1^{-1}$ .

By condition (iii) of the theorem and using exactly the same procedure as that used to obtain expression (17) we will get that expression which we will put in the following terms:

$$\hat{a}_1 \left[ \tau_1^{-1/2} \left( \hat{\tau}_1^{(\tau_1)} - \tau_1 \right) \right] + \sum_{i=1}^{q_1} \hat{a}_{i+1} \left[ \tau_1^{1/2} \left( \hat{\theta}_{1i}^{(\tau_1)} - \theta_{1i}^* \right) \right] - Z^{(\tau_1)} \xrightarrow{P} 0, \quad (28)$$

where

$$\begin{aligned}\hat{a}_1 &= \frac{1 - (1 - n/N)\pi_0^{(1)}(\hat{\boldsymbol{\theta}}_1^{(\tau_1)})}{(1 - n/N)\pi_0^{(1)}(\boldsymbol{\theta}_1^*)}, \quad \hat{a}_{i+1} = -\frac{1}{\pi_0^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_0^{(1)}(\boldsymbol{\theta}_{10}^{(\tau_1)})}{\partial \theta_i^{(1)}}, \quad i = 1, \dots, q_1, \\ Z^{(\tau_1)} &= \tau_1^{-1/2} \frac{\left( M^{(\tau_1)} + R_1^{(\tau_1)} \right) - \tau_1 \left[ 1 - (1 - n/N)\pi_0^{(1)}(\boldsymbol{\theta}_1^*) \right]}{(1 - n/N)\pi_0^{(1)}(\boldsymbol{\theta}_1^*)}, \quad (29)\end{aligned}$$

and  $\theta_{10}^{(\tau_1)}$  is between  $\hat{\theta}_1^{(\tau_1)}$  and  $\theta_1^*$ . Notice that conditions **(3)**-**(4)** and condition **(i)** of the theorem imply that  $\hat{a}_i \xrightarrow{P} a_i$ ,  $i = 1, \dots, q_1 + 1$ , where  $a_1 = \left[1 - (1 - n/N)\pi_0^{(1)}(\theta_1^*)\right] / (1 - n/N)\pi_0^{(1)}(\theta_1^*)$ , and  $a_{i+1} = -\left[\partial\pi_0^{(1)}(\theta_1^*)/\partial\theta_i^{(1)}\right] / \pi_0^{(1)}(\theta_1^*)$ ,  $i = 1, \dots, q_1$ .

Let  $\mathbf{Z}^{(\tau_1)} = \left[Z_1^{(\tau_1)}, Z_2^{(\tau_1)}, \dots, Z_{q_1}^{(\tau_1)}\right]'$ , then by the previous results we have that

$$\hat{\Psi}_1^{-1} \left[ \tau_1^{1/2} \left( \hat{\theta}_1^{(\tau_1)} - \theta_1^* \right)' \right] - \mathbf{Z}^{(\tau_1)} \xrightarrow{P} \mathbf{0}', \quad (30)$$

where  $\hat{\Psi}_1^{-1}$  is the  $q_1 \times q_1$  matrix whose elements are defined in (27).

We will show that  $\mathbf{Z}^{(\tau_1)} \xrightarrow{D} \mathbf{Z} \sim N_{q_1}(\mathbf{0}', \Psi_1^{-1})$  as  $\tau_1 \rightarrow \infty$ , where  $\mathbf{Z} = (Z_1, \dots, Z_{q_1})'$ , and that  $Z^{(\tau_1)} \xrightarrow{D} Z \sim N(0, a_1)$ , where  $Z^{(\tau_1)}$  is given by (29). To do this, we will associate with each element  $t \in U_1$ ,  $t = 1, \dots, \tau_1$ , a random vector  $\mathbf{V}_t^{(1)} = [V_{t,1}^{(1)}, \dots, V_{t,q_1}^{(1)}]'$  and a random variable  $V_t^{(1)}$  such that

- (a)  $V_{t,j}^{(1)} = [\tilde{\pi}_{\mathbf{x}}^{(1)}(\theta_1^*)]^{-1} \partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\theta_1^*) / \partial \theta_j^{(1)}$ ,  $j = 1, \dots, q_1$ , and  $V_t^{(1)} = 1$ , if  $t \in U_1 - S_0$  and its associated vector  $\mathbf{X}_t^{(1)}$  of link-indicator variables equals the vector  $\mathbf{x} \in \Omega - \{\mathbf{0}\}$ ;
- (b)  $V_{t,j}^{(1)} = 0$ ,  $j = 1, \dots, q_1$ , and  $V_t^{(1)} = -\left[1 - (1 - n/N)\pi_0^{(1)}(\theta_1^*)\right] / \left[(1 - n/N)\pi_0^{(1)}(\theta_1^*)\right]$ , if  $t \in U_1 - S_0$  and its associated vector  $\mathbf{X}_t^{(1)}$  of link-indicator variables equals the vector  $\mathbf{0} \in \Omega$ , and
- (c)  $V_{t,j}^{(1)} = [\pi_{\mathbf{x}}^{(A_l)}(\theta_1^*)]^{-1} \partial \pi_{\mathbf{x}}^{(A_l)}(\theta_1^*) / \partial \theta_j^{(1)}$ ,  $j = 1, \dots, q_1$ , and  $V_t^{(1)} = 1$ , if  $t \in A_l \in S_A$  and its associated vector  $\mathbf{X}_t^{(1)}$  of link-indicator variables equals the vector  $\mathbf{x} \in \Omega_{-l}$ .

Since

$$\begin{aligned} \tau_1^{-1/2} \sum_{t=1}^{\tau_1} V_{t,j}^{(1)} &= \tau_1^{-1/2} \left[ \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{R_{\mathbf{x}}^{(\tau_1)}}{\tilde{\pi}_{\mathbf{x}}^{(1)}(\theta_1^*)} \frac{\partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\theta_1^*)}{\partial \theta_j^{(1)}} + \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \frac{R_{\mathbf{x}}^{(A_l, \tau_1)}}{\pi_{\mathbf{x}}^{(A_l)}(\theta_1^*)} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\theta_1^*)}{\partial \theta_j^{(1)}} \right] \\ &= Z_j^{(\tau_1)}, \quad j = 1, \dots, q_1, \end{aligned}$$

it follows that  $\mathbf{Z}^{(\tau_1)} = \tau_1^{-1/2} \sum_{t=1}^{\tau_1} \mathbf{V}_t^{(1)}$ , and

$$\begin{aligned} \tau_1^{-1/2} \sum_{t=1}^{\tau_1} V_t^{(1)} &= \tau_1^{-1/2} \left[ M^{(\tau_1)} + R_1^{(\tau_1)} - \left( \tau_1 - M^{(\tau_1)} - R_1^{(\tau_1)} \right) \frac{1 - (1 - n/N)\pi_0^{(1)}(\theta_1^*)}{(1 - n/N)\pi_0^{(1)}(\theta_1^*)} \right] \\ &= Z^{(\tau_1)}. \end{aligned}$$

From the definition of  $V_{t,j}^{(1)}$  and  $V_t^{(1)}$  we have that

$$\begin{aligned}\Pr \left\{ V_{t,j}^{(1)} = [\tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)]^{-1} \partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*) / \partial \theta_j^{(1)} \right\} &= (1 - n/N) \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*), \mathbf{x} \in \Omega - \{\mathbf{0}\}, \\ & j = 1, \dots, q_1, \\ \Pr \left\{ V_{t,j}^{(1)} = 0 \right\} &= (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*), \quad j = 1, \dots, q_1, \\ \Pr \left\{ V_{t,j}^{(1)} = [\pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)]^{-1} \partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*) / \partial \theta_j^{(1)} \right\} &= (1/N) \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*), \mathbf{x} \in \Omega_{-l}, j = 1, \dots, q_1, \\ & l = 1, \dots, n,\end{aligned}$$

and

$$\begin{aligned}\Pr \left\{ V_t^{(1)} = 1 \right\} &= (1 - n/N) \left[ 1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right] + n/N \text{ and} \\ \Pr \left\{ V_t^{(1)} = - \left[ 1 - (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right] / \left[ (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right] \right\} &= (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*); \\ \text{therefore, the expected values of the variables } V_{t,j}^{(1)} \text{ and } V_t^{(1)} &\text{ are}\end{aligned}$$

$$\begin{aligned}E \left( V_{t,j}^{(1)} \right) &= \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*) / \partial \theta_j^{(1)} (1 - n/N) \left[ 1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right] \\ &+ \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*) / \partial \theta_j^{(1)} (1/N) = 0, \quad j = 1, \dots, q_1,\end{aligned}$$

and

$$E \left( V_t^{(1)} \right) = (1 - n/N) \left[ 1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right] + n/N - \left[ 1 - (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right] = 0$$

because of (22). Thus,  $E \left( \mathbf{V}_t^{(1)} \right) = \mathbf{0}$  and  $E \left( V_t^{(1)} \right) = 0$ ,  $t = 1, \dots, \tau_1$ . Furthermore, their variances are

$$\begin{aligned}V \left( V_{t,j}^{(1)} \right) &= (1 - n/N) \left[ 1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right] \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{1}{\tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)} \left[ \frac{\partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \right]^2 \\ &+ \frac{1}{N} \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \frac{1}{\pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)} \left[ \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \right]^2, \quad j = 1, \dots, q_1,\end{aligned}$$

and

$$\begin{aligned}V \left( V_t^{(1)} \right) &= (1 - n/N) \left[ 1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right] + n/N + \frac{\left[ 1 - (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right]^2}{(1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)} \\ &= \frac{1 - (1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{(1 - n/N) \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)},\end{aligned}$$



and their covariances are

$$\begin{aligned} Cov \left( V_{t,j}^{(1)}, V_{t,j'}^{(1)} \right) = & \left( 1 - \frac{n}{N} \right) \left[ 1 - \pi_0^{(1)}(\theta_1^*) \right] \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{1}{\tilde{\pi}_{\mathbf{x}}^{(1)}(\theta_1^*)} \frac{\partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\theta_1^*)}{\partial \theta_j^{(1)}} \frac{\partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\theta_1^*)}{\partial \theta_{j'}^{(1)}} \\ & + \frac{1}{N} \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \frac{1}{\partial \pi_{\mathbf{x}}^{(A_l)}(\theta_1^*)} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\theta_1^*)}{\partial \theta_j^{(1)}} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\theta_1^*)}{\partial \theta_{j'}^{(1)}}, \quad j, j' = 1, \dots, q_1, \quad j \neq j', \end{aligned}$$

and

$$Cov \left( V_t^{(1)}, V_{t,j}^{(1)} \right) = \left( 1 - \frac{n}{N} \right) \left[ 1 - \pi_0^{(1)}(\theta_1^*) \right] \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{\partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\theta_1^*)}{\partial \theta_j^{(1)}} + \frac{1}{N} \sum_{l=1}^n \sum_{\mathbf{x} \in \Omega_{-l}} \frac{\partial \pi_{\mathbf{x}}^{(A_l)}(\theta_1^*)}{\partial \theta_j^{(1)}} = 0.$$

Therefore, the variance-covariance matrix of  $\mathbf{V}_t^{(1)}$  is  $\Psi_1^{-1}$ .

Finally, since the  $(\mathbf{V}_t^{(1)'}, V_t^{(1)})'$ ,  $t = 1, \dots, \tau_1$ , are independent and identically distributed random vectors, by the central limit theorem it follows that

$$(\mathbf{Z}^{(\tau_1)'}, Z^{(\tau_1)})' = \tau_1^{-1/2} \sum_{t=1}^{\tau_1} (\mathbf{V}_t^{(1)'}, V_t^{(1)})' \xrightarrow{D} (\mathbf{Z}', Z)' \sim N_{q_1+1} \left( \mathbf{0}'_{q_1+1}, \begin{bmatrix} \Psi_1^{-1} & \mathbf{0}' \\ \mathbf{0} & a_1 \end{bmatrix} \right).$$

Thus,  $\mathbf{Z}^{(\tau_1)} \xrightarrow{D} \mathbf{Z} \sim N_{q_1}(\mathbf{0}', \Psi_1^{-1})$  and  $Z^{(\tau_1)} \xrightarrow{D} Z \sim N(0, a_1)$ . Consequently by (30)

$$\tau_1^{1/2} \left[ \hat{\theta}_1^{(\tau_1)} - \theta_1^* \right] \xrightarrow{D} (\Psi_1 \mathbf{Z})' \sim N_{q_1}(\mathbf{0}, \Psi_1)$$

as  $\hat{\Psi}_1 \xrightarrow{P} \Psi_1$ .

At last, from (28) and the previous results

$$\begin{aligned} \tau_1^{-1/2} \left( \hat{\tau}_1^{(\tau_1)} - \tau_1 \right) & \xrightarrow{D} \frac{1}{a_1} \left\{ Z - \sum_{i=1}^{q_1} a_{i+1} [\Psi_1 \mathbf{Z}]_i \right\} \\ & = \frac{(1 - n/N) \pi_0^{(1)}(\theta_1^*)}{1 - (1 - n/N) \pi_0^{(1)}(\theta_1^*)} \left[ Z + \frac{1}{\pi_0^{(1)}(\theta_1^*)} \left[ \nabla \pi_0^{(1)}(\theta_1^*) \right]' \Psi_1 \mathbf{Z} \right] \sim N(0, \sigma^2), \end{aligned}$$

where  $[\Psi_1 \mathbf{Z}]_i$  is the  $i$ -th element of  $\Psi_1 \mathbf{Z}$  and

$$\sigma^2 = \frac{1 - n/N}{1 - (1 - n/N) \pi_0^{(1)}(\theta_1^*)} \left\{ \pi_0^{(1)}(\theta_1^*) + \frac{(1 - n/N) \left[ \nabla \pi_0^{(1)}(\theta_1^*) \right]' \Psi_1 \left[ \nabla \pi_0^{(1)}(\theta_1^*) \right]}{1 - (1 - n/N) \pi_0^{(1)}(\theta_1^*)} \right\}.$$

□

#### 4.4 Consistency of the UMLE and CMLE of $(\tau_1, \theta_1^*)$

To prove the consistency of the UMLE and CMLE we will use condition **(5)** and the following inequality of information theory: If  $\sum a_i$  and  $\sum b_i$  are convergent series of positive numbers such that  $\sum a_i \geq \sum b_i$ , then  $\sum a_i \log(b_i) \leq \sum a_i \log(a_i)$ , and the equality is attained if and only if  $a_i = b_i$ . See Rao (1973, p. 58).

#### 4.4.1 Consistency of the UMLE

Let us first consider  $\hat{\boldsymbol{\theta}}_1^{(U)}$ . Using (3) and (6) and the definition of the UMLE  $(\hat{\tau}_1^{(U)}, \hat{\boldsymbol{\theta}}_1^{(U)})$  we get that

$$\begin{aligned} l_{(1)}(\hat{\tau}_1^{(U)}, \hat{\boldsymbol{\theta}}_1^{(U)}) &= \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} R_{\mathbf{x}}^{(\tau_1)} \ln \left\{ \pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(U)}) / \left[ 1 - \pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(U)}) \right] \right\} \\ &+ \sum_{i=1}^n \sum_{\mathbf{x} \in \Omega_{-i}} R_{\mathbf{x}}^{(A_i, \tau_1)} \ln \left[ \pi_{\mathbf{x}}^{(A_i)}(\hat{\boldsymbol{\theta}}_1^{(U)}) \right] + \ln \left[ L_{MULT}(\hat{\tau}_1^{(U)}) \right] + \ln \left[ L_{12}(\hat{\tau}_1^{(U)}, \hat{\boldsymbol{\theta}}_1^{(U)}) \right] + C \\ &\geq \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} R_{\mathbf{x}}^{(\tau_1)} \ln \left\{ \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*) / \left[ 1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right] \right\} + \sum_{i=1}^n \sum_{\mathbf{x} \in \Omega_{-i}} R_{\mathbf{x}}^{(A_i, \tau_1)} \ln \left[ \pi_{\mathbf{x}}^{(A_i)}(\boldsymbol{\theta}_1^*) \right] \\ &+ \ln \left[ L_{MULT}(\tau_1) \right] + \ln \left[ L_{12}(\tau_1, \boldsymbol{\theta}_1^*) \right] + C = l_{(1)}(\tau_1, \boldsymbol{\theta}_1^*), \end{aligned}$$

where  $C$  depends only on observable variables. Since  $\ln \left[ L_{MULT}(\hat{\tau}_1^{(U)}) \right]$  and  $\ln \left[ L_{12}(\hat{\tau}_1^{(U)}, \hat{\boldsymbol{\theta}}_1^{(U)}) \right]$  are nonpositive we have that

$$\begin{aligned} &\sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{R_{\mathbf{x}}^{(\tau_1)}}{R_1^{(\tau_1)}} \ln \left[ \frac{\pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(U)})}{1 - \pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(U)})} \right] + \sum_{i=1}^n \frac{M_i^{(\tau_1)}}{R_1^{(\tau_1)}} \sum_{\mathbf{x} \in \Omega_{-i}} \frac{R_{\mathbf{x}}^{(A_i, \tau_1)}}{M_i^{(\tau_1)}} \ln \left[ \pi_{\mathbf{x}}^{(A_i)}(\hat{\boldsymbol{\theta}}_1^{(U)}) \right] \\ &\geq \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{R_{\mathbf{x}}^{(\tau_1)}}{R_1^{(\tau_1)}} \ln \left[ \frac{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)} \right] + \sum_{i=1}^n \frac{M_i^{(\tau_1)}}{R_1^{(\tau_1)}} \sum_{\mathbf{x} \in \Omega_{-i}} \frac{R_{\mathbf{x}}^{(A_i, \tau_1)}}{M_i^{(\tau_1)}} \ln \left[ \pi_{\mathbf{x}}^{(A_i)}(\boldsymbol{\theta}_1^*) \right] \\ &+ \ln \left[ L_{MULT}(\tau_1) \right] / R_1^{(\tau_1)} + \ln \left[ L_{12}(\tau_1, \boldsymbol{\theta}_1^*) \right] / R_1^{(\tau_1)}. \end{aligned} \quad (31)$$

Now, since

$$1 = \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{R_{\mathbf{x}}^{(\tau_1)}}{R_1^{(\tau_1)}} = \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{\pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(U)})}{1 - \pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(U)})} \quad \text{and} \quad 1 = \sum_{\mathbf{x} \in \Omega_{-i}} \frac{R_{\mathbf{x}}^{(A_i, \tau_1)}}{M_i^{(\tau_1)}} = \sum_{\mathbf{x} \in \Omega_{-i}} \pi_{\mathbf{x}}^{(A_i)}(\hat{\boldsymbol{\theta}}_1^{(U)}),$$

$i = 1, \dots, n,$

using  $n + 1$  times the previously indicated information theory inequality we have that

$$\begin{aligned} &\sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{R_{\mathbf{x}}^{(\tau_1)}}{R_1^{(\tau_1)}} \ln \left[ R_{\mathbf{x}}^{(\tau_1)} / R_1^{(\tau_1)} \right] + \sum_{i=1}^n \frac{M_i^{(\tau_1)}}{R_1^{(\tau_1)}} \sum_{\mathbf{x} \in \Omega_{-i}} \frac{R_{\mathbf{x}}^{(A_i, \tau_1)}}{M_i^{(\tau_1)}} \ln \left[ R_{\mathbf{x}}^{(A_i, \tau_1)} / M_i^{(\tau_1)} \right] \\ &\geq \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{R_{\mathbf{x}}^{(\tau_1)}}{R_1^{(\tau_1)}} \ln \left[ \frac{\pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(U)})}{1 - \pi_{\mathbf{0}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(U)})} \right] + \sum_{i=1}^n \frac{M_i^{(\tau_1)}}{R_1^{(\tau_1)}} \sum_{\mathbf{x} \in \Omega_{-i}} \frac{R_{\mathbf{x}}^{(A_i, \tau_1)}}{M_i^{(\tau_1)}} \ln \left[ \pi_{\mathbf{x}}^{(A_i)}(\hat{\boldsymbol{\theta}}_1^{(U)}) \right]. \end{aligned} \quad (32)$$

Thus, by (31) and (32) we get that

$$\begin{aligned}
& \geq \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{R_{\mathbf{x}}^{(\tau_1)}}{R_1^{(\tau_1)}} \ln \left\{ \frac{\pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(U)}) / [1 - \pi_0^{(1)}(\hat{\boldsymbol{\theta}}_1^{(U)})]}{R_{\mathbf{x}}^{(1)} / R_1^{(\tau_1)}} \right\} \\
& \quad + \sum_{i=1}^n \frac{M_i^{(\tau_1)}}{R_1^{(\tau_1)}} \sum_{\mathbf{x} \in \Omega_{-i}} \frac{R_{\mathbf{x}}^{(A_i, \tau_1)}}{M_i^{(\tau_1)}} \ln \left[ \frac{\pi_{\mathbf{x}}^{(A_i)}(\hat{\boldsymbol{\theta}}_1^{(U)})}{R_{\mathbf{x}}^{(A_i, \tau_1)} / M_i^{(\tau_1)}} \right] \\
& \geq \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{R_{\mathbf{x}}^{(\tau_1)}}{R_1^{(\tau_1)}} \ln \left\{ \frac{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*) / [1 - \pi_0^{(1)}(\boldsymbol{\theta}_1^*)]}{R_{\mathbf{x}}^{(1)} / R_1^{(\tau_1)}} \right\} + \sum_{i=1}^n \frac{M_i^{(\tau_1)}}{R_1^{(\tau_1)}} \sum_{\mathbf{x} \in \Omega_{-i}} \frac{R_{\mathbf{x}}^{(A_i, \tau_1)}}{M_i^{(\tau_1)}} \\
& \quad \times \ln \left[ \frac{\pi_{\mathbf{x}}^{(A_i)}(\boldsymbol{\theta}_1^*)}{R_{\mathbf{x}}^{(A_i, \tau_1)} / M_i^{(\tau_1)}} \right] + \ln [L_{MULT}(\tau_1)] / R_1^{(\tau_1)} + \ln [L_{12}(\tau_1, \boldsymbol{\theta}_1^*)] / R_1^{(\tau_1)}. \quad (33)
\end{aligned}$$

From the unconditional distributions of  $M_i^{(\tau_1)}$ ,  $M^{(\tau_1)}$  and  $R_1^{(\tau_1)}$ ,  $R_{\mathbf{x}}^{(A_i, \tau_1)}$  and  $R_{\mathbf{x}}^{(A_i, \tau_1)}$  indicated in Subsection 4.1, it follows that  $R_{\mathbf{x}}^{(\tau_1)} / R_1^{(\tau_1)} \xrightarrow{P} \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*) / [1 - \pi_0^{(1)}(\boldsymbol{\theta}_1^*)]$ ,  $R_{\mathbf{x}}^{(A_i, \tau_1)} / M_i^{(\tau_1)} \xrightarrow{P} \pi_{\mathbf{x}}^{(A_i)}(\boldsymbol{\theta}_1^*)$  and  $M_i^{(\tau_1)} / R_1^{(\tau_1)} \xrightarrow{P} 1 / \{(N - n)[1 - \pi_0^{(1)}(\boldsymbol{\theta}_1^*)]\}$ . Therefore, the first two summands of the last term of the double inequality (33) converges to zero in probability, In addition, since  $R_1^{(\tau_1)} / \tau_1 \xrightarrow{P} 1 - \pi_0^{(1)}(\boldsymbol{\theta}_1^*)$ , and from well known results of large deviations theory (see Varadhan, 2008), we have that for the binomial probability  $L_{12}(\tau_1, \boldsymbol{\theta}_1^*)$ :

$$\begin{aligned}
\frac{\ln [L_{12}(\tau_1, \boldsymbol{\theta}_1^*)]}{R_1^{(\tau_1)}} &= -\frac{\tau_1 - M^{(\tau_1)}}{R_1^{(\tau_1)}} \left\{ \frac{R_1^{(\tau_1)}}{\tau_1 - M^{(\tau_1)}} \ln \left[ \frac{R_1^{(\tau_1)} / (\tau_1 - M^{(\tau_1)})}{1 - \pi_0^{(1)}(\boldsymbol{\theta}_1^*)} \right] + \frac{\tau_1 - R_1^{(\tau_1)}}{\tau_1 - M^{(\tau_1)}} \right. \\
&\quad \times \ln \left[ \frac{(\tau_1 - R_1^{(\tau_1)}) / (\tau_1 - M^{(\tau_1)})}{\pi_0^{(1)}(\boldsymbol{\theta}_1^*)} \right] \left. \right\} + \frac{\tau_1 - M^{(\tau_1)}}{R_1^{(\tau_1)}} o_p(1) \\
&\xrightarrow{P} -\frac{[1 - \pi_0^{(1)}(\boldsymbol{\theta}_1^*)] \ln(1) + \pi_0^{(1)}(\boldsymbol{\theta}_1^*) \ln(1)}{1 - \pi_0^{(1)}(\boldsymbol{\theta}_1^*)} = 0,
\end{aligned}$$

and for the multinomial probability  $L_{MULT}(\tau_1)$ :

$$\begin{aligned}
\frac{\ln [L_{MULT}(\tau_1)]}{R_1^{(\tau_1)}} &= -\frac{\tau_1}{R_1^{(\tau_1)}} \left\{ \sum_{i=1}^n \frac{M_i^{(\tau_1)}}{\tau_1} \ln \left[ \frac{M_i^{(\tau_1)} / \tau_1}{1/N} \right] + \frac{\tau_1 - M^{(\tau_1)}}{\tau_1} \ln \left[ \frac{(\tau_1 - M^{(\tau_1)}) / \tau_1}{1 - n/N} \right] \right\} \\
&\quad + \frac{\tau_1}{R_1^{(\tau_1)}} o_p(1) \xrightarrow{P} -\left\{ \sum_{i=1}^n \frac{1}{N} \ln(1) + (1 - n/N) \ln(1) \right\} / \left\{ [1 - \pi_0^{(1)}(\boldsymbol{\theta}_1^*)] (1 - n/N) \right\} = 0.
\end{aligned}$$

The previous results imply that the last term of the double inequality (33) converges to zero in probability, and consequently so does the middle term.

Thus,

$$\begin{aligned}
& \sum_{\mathbf{x} \in \Omega - \{0\}} \frac{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{1 - \pi_0^{(1)}(\boldsymbol{\theta}_1^*)} \ln \left\{ \frac{\pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(U)}) / [1 - \pi_0^{(1)}(\hat{\boldsymbol{\theta}}_1^{(U)})]}{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*) / [1 - \pi_0^{(1)}(\boldsymbol{\theta}_1^*)]} \right\} \\
& + \frac{1}{(N - n) [1 - \pi_0^{(1)}(\boldsymbol{\theta}_1^*)]} \sum_{i=1}^n \sum_{\mathbf{x} \in \Omega_{-i}} \pi_{\mathbf{x}}^{(A_i)}(\boldsymbol{\theta}_1^*) \ln \left[ \frac{\pi_{\mathbf{x}}^{(A_i)}(\hat{\boldsymbol{\theta}}_1^{(U)})}{\pi_{\mathbf{x}}^{(A_i)}(\boldsymbol{\theta}_1^*)} \right] \\
& = \sum_{\mathbf{x} \in \Omega - \{0\}} \frac{R_{\mathbf{x}}^{(\tau_1)}}{R_1^{(\tau_1)}} \ln \left\{ \frac{\pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(U)}) / [1 - \pi_0^{(1)}(\hat{\boldsymbol{\theta}}_1^{(U)})]}{R_{\mathbf{x}}^{(\tau_1)} / R_1^{(\tau_1)}} \right\} \\
& + \sum_{i=1}^n \frac{M_i^{(\tau_1)}}{R_1^{(\tau_1)}} \sum_{\mathbf{x} \in \Omega_{-i}} \frac{R_{\mathbf{x}}^{(A_i, \tau_1)}}{M_i^{(\tau_1)}} \ln \left[ \frac{\pi_{\mathbf{x}}^{(A_i)}(\hat{\boldsymbol{\theta}}_1^{(U)})}{R_{\mathbf{x}}^{(A_i, \tau_1)} / M_i^{(\tau_1)}} \right] \\
& + \sum_{\mathbf{x} \in \Omega - \{0\}} \left[ \frac{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{1 - \pi_0^{(1)}(\boldsymbol{\theta}_1^*)} - \frac{R_{\mathbf{x}}^{(\tau_1)}}{R_1^{(\tau_1)}} \right] \ln \left\{ \frac{\pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(U)}) / [1 - \pi_0^{(1)}(\hat{\boldsymbol{\theta}}_1^{(U)})]}{R_{\mathbf{x}}^{(\tau_1)} / R_1^{(\tau_1)}} \right\} \\
& + \sum_{i=1}^n \sum_{\mathbf{x} \in \Omega_{-i}} \left[ \frac{\pi_{\mathbf{x}}^{(A_i)}(\boldsymbol{\theta}_1^*)}{(N - n) [1 - \pi_0^{(1)}(\boldsymbol{\theta}_1^*)]} - \frac{R_{\mathbf{x}}^{(A_i)}}{R_1^{(\tau_1)}} \right] \ln \left[ \frac{\pi_{\mathbf{x}}^{(A_i)}(\hat{\boldsymbol{\theta}}_1^{(U)})}{R_{\mathbf{x}}^{(A_i, \tau_1)} / M_i^{(\tau_1)}} \right] \\
& + \sum_{\mathbf{x} \in \Omega - \{0\}} \frac{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{1 - \pi_0^{(1)}(\boldsymbol{\theta}_1^*)} \ln \left\{ \frac{R_{\mathbf{x}}^{(\tau_1)} / R_1^{(\tau_1)}}{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*) / [1 - \pi_0^{(1)}(\boldsymbol{\theta}_1^*)]} \right\} \\
& + \frac{1}{(N - n) [1 - \pi_0^{(1)}(\boldsymbol{\theta}_1^*)]} \sum_{i=1}^n \sum_{\mathbf{x} \in \Omega_{-i}} \pi_{\mathbf{x}}^{(A_i)}(\boldsymbol{\theta}_1^*) \ln \left[ \frac{R_{\mathbf{x}}^{(A_i, \tau_1)} / M_i^{(\tau_1)}}{\pi_{\mathbf{x}}^{(A_i)}(\boldsymbol{\theta}_1^*)} \right] \xrightarrow{P} 0
\end{aligned}$$

as  $\ln \left\{ \left[ \pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(U)}) / (1 - \pi_0^{(1)}(\hat{\boldsymbol{\theta}}_1^{(U)})) \right] / [R_{\mathbf{x}}^{(\tau_1)} / R_1^{(\tau_1)}] \right\}$  and  $\ln \left[ \pi_{\mathbf{x}}^{(A_i)}(\hat{\boldsymbol{\theta}}_1^{(U)}) / (R_{\mathbf{x}}^{(A_i, \tau_1)} / M_i^{(\tau_1)}) \right]$  are bounded as  $\tau_1 \rightarrow \infty$  (otherwise the middle term of the inequality (33) would not converge to zero). Finally, condition (5) implies that for any  $\delta_1 > 0$  we have that  $\Pr \left\{ \left\| \hat{\boldsymbol{\theta}}_1^{(U)} - \boldsymbol{\theta}_1^* \right\| \leq \delta_1 \right\} \rightarrow 1$ , that is,  $\hat{\boldsymbol{\theta}}_1^{(U)} \xrightarrow{P} \boldsymbol{\theta}_1^*$ .

Straightforward results of the previous one are the following:  $\pi_{\mathbf{x}}^{(1)}(\hat{\boldsymbol{\theta}}_1^{(U)}) \xrightarrow{P} \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)$ ,  $\mathbf{x} \in \Omega$ , and  $\pi_{\mathbf{x}}^{(A_i)}(\hat{\boldsymbol{\theta}}_1^{(U)}) \xrightarrow{P} \pi_{\mathbf{x}}^{(A_i)}(\boldsymbol{\theta}_1^*)$ ,  $\mathbf{x} \in \Omega_{-i}$ ,  $i = 1, \dots, n$ , as  $\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1)$  and  $\pi_{\mathbf{x}}^{(A_i)}(\boldsymbol{\theta}_1)$  are assumed to be continuous functions of  $\boldsymbol{\theta}_1$ .

With respect to  $\hat{\tau}_1^{(U)}$ , from expression (5) we have that the difference between  $\hat{\tau}_1^{(U)}$  and  $(M^{(\tau_1)} + R_1^{(\tau_1)}) / [1 - (1 - n/N)\pi_0^{(1)}(\hat{\boldsymbol{\theta}}_1^{(U)})]$  is less than 1. Thus,  $\left\{ \hat{\tau}_1^{(U)} - (M^{(\tau_1)} + R_1^{(\tau_1)}) / [1 - (1 - n/N)\pi_0^{(1)}(\hat{\boldsymbol{\theta}}_1^{(U)})] \right\} / \tau_1 = \hat{\tau}_1^{(U)} / \tau_1 - [(M^{(\tau_1)} + R_1^{(\tau_1)}) / \tau_1] / [1 - (1 - n/N)]$

$\pi_0^{(1)} \left( \hat{\theta}_1^{(U)} \right) \Big] \xrightarrow{P} 0$ , and since the second term of the last difference converges to 1 in probability so does  $\hat{\tau}_1^{(U)} / \tau_1$ .

#### 4.4.2 Consistency of the CMLE

By the definition of the CMLE  $\hat{\theta}_1^{(C)}$ , we have that

$$\begin{aligned} \frac{\ln[L_{11}(\hat{\theta}_1^{(C)})L_0(\hat{\theta}_1^{(C)})]}{R_1^{(\tau_1)}} &= \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{R_{\mathbf{x}}^{(\tau_1)}}{R_1^{(\tau_1)}} \ln \left[ \frac{\pi_{\mathbf{x}}^{(1)} \left( \hat{\theta}_1^{(C)} \right)}{1 - \pi_{\mathbf{0}}^{(1)} \left( \hat{\theta}_1^{(C)} \right)} \right] \\ &\quad + \sum_{i=1}^n \frac{M_i^{(\tau_1)}}{R_1^{(\tau_1)}} \sum_{\mathbf{x} \in \Omega_{-i}} \frac{R_{\mathbf{x}}^{(A_i, \tau_1)}}{M_i^{(\tau_1)}} \ln \left[ \pi_{\mathbf{x}}^{(A_i)} \left( \hat{\theta}_1^{(C)} \right) \right] + C \\ &\geq \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{R_{\mathbf{x}}^{(\tau_1)}}{R_1^{(\tau_1)}} \ln \left[ \frac{\pi_{\mathbf{x}}^{(1)} (\theta_1^*)}{1 - \pi_{\mathbf{0}}^{(1)} (\theta_1^*)} \right] \\ &\quad + \sum_{i=1}^n \frac{M_i^{(\tau_1)}}{R_1^{(\tau_1)}} \sum_{\mathbf{x} \in \Omega_{-i}} \frac{R_{\mathbf{x}}^{(A_i, \tau_1)}}{M_i^{(\tau_1)}} \ln \left[ \pi_{\mathbf{x}}^{(A_i)} (\theta_1^*) \right] + C \\ &= \frac{\ln[L_{11}(\theta_1^*)L_0(\theta_1^*)]}{R_1^{(\tau_1)}}, \end{aligned}$$

where  $C$  depends only on observable variables.

Using the same procedure as that used in the case of the UMLE  $\hat{\theta}_1^{(U)}$  we will get the double inequality (33) but in terms of  $\hat{\theta}_1^{(C)}$  instead of  $\hat{\theta}_1^{(U)}$  and without the terms  $\ln[L_{MULT}(\tau_1)]/R_1^{(\tau_1)}$  and  $\ln[L_{12}(\tau_1, \theta_1^*)]/R_1^{(\tau_1)}$ . Consequently, we will also have that  $\hat{\theta}_1^{(C)} \xrightarrow{P} \theta_1^*$ ,  $\pi_{\mathbf{x}}^{(1)} \left( \hat{\theta}_1^{(C)} \right) \xrightarrow{P} \pi_{\mathbf{x}}^{(1)}(\theta_1^*)$ ,  $\mathbf{x} \in \Omega$ ,  $\pi_{\mathbf{x}}^{(A_i)} \left( \hat{\theta}_1^{(C)} \right) \xrightarrow{P} \pi_{\mathbf{x}}^{(A_i)}(\theta_1^*)$ ,  $\mathbf{x} \in \Omega_{-i}$ ,  $i = 1, \dots, n$ , and  $\hat{\tau}_1^{(C)} / \tau_1 \xrightarrow{P} 1$ , where the last result is obtained by using expression (8) and the same arguments as those used to prove that  $\hat{\tau}_1^{(U)} / \tau_1 \xrightarrow{P} 1$ .

### 4.5 Asymptotic distributions of the UMLEs and CMLEs of $\tau_1$ and $\theta_1^*$

#### 4.5.1 Asymptotic multivariate normal distribution of the UMLE of $(\tau_1, \theta_1^*)$

We will prove the asymptotic multivariate normal distribution of  $\left( \tau_1^{-1/2} \hat{\tau}_1^{(U)}, \tau_1^{1/2} \hat{\theta}_1^{(U)} \right)$  by proving that this estimator satisfies the conditions of Theorem 1. Condition (i) was already proved in the previous section. From expression (5) it follows that  $\left( \hat{\tau}_1^{(U)}, \hat{\theta}_1^{(U)} \right)$  satisfies condition (ii). Finally, by the definition of the UMLEs we have that condition (iii) is also satisfied. Thus, by Theorem 1,  $\left[ \tau_1^{-1/2} \left( \hat{\tau}_1^{(U)} - \tau_1 \right), \tau_1^{1/2} \left( \hat{\theta}_1^{(U)} - \theta_1^* \right) \right] \xrightarrow{D} N_{q_1+1}(\mathbf{0}, \Sigma_1)$ . This result

implies that  $\tau_1^{-1/2} \left( \hat{\tau}_1^{(U)} - \tau_1 \right) \xrightarrow{D} N(0, \sigma_{1U}^2)$  and  $\tau_1^{1/2} \left( \hat{\theta}_1^{(U)} - \theta_1^* \right) \xrightarrow{D} N_{q_1}(\mathbf{0}, \Sigma_{122})$ , where

$$\sigma_{1U}^2 = \frac{1 - n/N}{1 - (1 - n/N)\pi_0^{(1)}(\theta_1^*)} \left\{ \pi_0^{(1)}(\theta_1^*) + \frac{1 - n/N}{1 - (1 - n/N)\pi_0^{(1)}(\theta_1^*)} \left[ \nabla \pi_0^{(1)}(\theta_1^*) \right]' \right. \\ \left. \times \Sigma_{122} \left[ \nabla \pi_0^{(1)}(\theta_1^*) \right] \right\}, \quad (34)$$

$$\Sigma_{122} = \left\{ \Sigma_{122}^{-1} - \frac{1 - n/N}{\pi_0^{(1)}(\theta_1^*) \left[ 1 - (1 - n/N)\pi_0^{(1)}(\theta_1^*) \right]} \left[ \nabla \pi_0^{(1)}(\theta_1^*) \right] \left[ \nabla \pi_0^{(1)}(\theta_1^*) \right]' \right\}^{-1}, \quad (35)$$

$\nabla \pi_0^{(1)}(\theta_1^*)$  is the gradient of  $\pi_0^{(1)}(\theta_1)$  evaluated at  $\theta_1^*$  and  $\Sigma_{122}^{-1}$  is the  $q_1 \times q_1$  submatrix of  $\Sigma_1^{-1}$  obtained by removing its first row and first column.

#### 4.5.2 Asymptotic multivariate normal distribution of the CMLE $\hat{\theta}_1^{(U)}$ and asymptotic normal distribution of the CMLE $\hat{\tau}_1^{(C)}$

The CMLE  $\left( \tau_1^{-1/2} \hat{\tau}_1^{(C)}, \tau_1^{1/2} \hat{\theta}_1^{(C)} \right)$  does not have an asymptotic multivariate normal distribution since this estimator does not satisfy condition (iii) of Theorem 1. To see this, notice that by (7) it follows that  $\partial \left\{ \ln \left[ L_{11} \left( \hat{\theta}_1^{(C)} \right) L_0 \left( \hat{\theta}_1^{(C)} \right) \right] \right\} / \partial \theta_j^{(1)} = 0$ . Therefore,

$$\tau_1^{-1/2} \frac{\partial}{\partial \theta_j^{(1)}} l_{(1)} \left( \hat{\tau}_1^{(C)}, \hat{\theta}_1^{(C)} \right) = \tau_1^{-1/2} \frac{\partial}{\partial \theta_j^{(1)}} \ln \left[ L_{12} \left( \hat{\tau}_1^{(C)}, \hat{\theta}_1^{(C)} \right) \right] \\ = \tau_1^{-1/2} \frac{\partial \pi_0^{(1)} \left( \hat{\theta}_1^{(C)} \right)}{\partial \theta_j^{(1)}} \left[ \frac{\hat{\tau}_1^{(C)} - M(\tau_1) - R_1^{(\tau_1)}}{\pi_0^{(1)} \left( \hat{\theta}_1^{(C)} \right)} - \frac{R_1^{(\tau_1)}}{1 - \pi_0^{(1)} \left( \hat{\theta}_1^{(C)} \right)} \right].$$

By using expression (8) and after some algebraic steps we get that

$$\tau_1^{-1/2} \frac{\partial}{\partial \theta_j^{(1)}} l_{(1)} \left( \hat{\tau}_1^{(C)}, \hat{\theta}_1^{(C)} \right) = \left[ \frac{\partial \pi_0^{(1)} \left( \hat{\theta}_1^{(C)} \right)}{\partial \theta_j^{(1)}} \right] \\ \times \left\{ \frac{\hat{\tau}_1^{(C)} - \left( M(\tau_1) + R_1^{(\tau_1)} \right) / \left[ 1 - (1 - n/N)\pi_0^{(1)} \left( \hat{\theta}_1^{(C)} \right) \right]}{\tau_1^{1/2} \pi_0^{(1)} \left( \hat{\theta}_1^{(C)} \right)} \right. \\ \left. + \tau_1^{1/2} \frac{\left( M(\tau_1) / \tau_1 \right) (1 - n/N) \left[ 1 - \pi_0^{(1)} \left( \hat{\theta}_1^{(C)} \right) \right] - \left( R_1^{(\tau_1)} / \tau_1 \right) (n/N)}{\left[ 1 - (1 - n/N)\pi_0^{(1)} \left( \hat{\theta}_1^{(C)} \right) \right] \left[ 1 - \pi_0^{(1)} \left( \hat{\theta}_1^{(C)} \right) \right]} \right\}. \quad (36)$$

From (8) and the fact that  $\pi_0^{(1)}(\hat{\theta}_1^{(C)}) \xrightarrow{P} \pi_0^{(1)}(\theta_1^*)$ , it follows that the order of magnitude of the first term in the curly brackets of (36) is  $O_p(\tau_1^{-1/2})$ . On the other hand, since  $M^{(\tau_1)}/\tau_1 = n/N + O_p(\tau_1^{-1/2})$ ,  $R_1^{(\tau_1)}/\tau_1 = (1 - n/N) \left[1 - \pi_0^{(1)}(\theta_1^*)\right] + O_p(\tau_1^{-1/2})$  and, as we will show in the next paragraph,  $\hat{\theta}_1^{(C)} = \theta_1^* + O_p(\tau_1^{-1/2})$ , it follows that the order of the second term in the curly brackets of (36) is  $O_p(1)$ ; therefore (36) does not converge to zero in probability.

Nevertheless, although  $\left[\tau_1^{-1/2}(\hat{\tau}_1^{(C)} - \tau_1), \tau_1^{1/2}(\hat{\theta}_1^{(C)} - \theta_1^*)\right]$  does not have an asymptotic multivariate normal distribution,  $\tau_1^{1/2}(\hat{\theta}_1^{(C)} - \theta_1^*)$  does have. To prove this, we will show that conditions (i) and (ii) of Theorem 2 are satisfied. In the previous section we proved that  $\hat{\theta}_1^{(C)}$  satisfies condition (i), and from (7) we have that  $\hat{\theta}_1^{(C)}$  satisfies condition (ii). Thus by Theorem 2,  $\tau_1^{1/2}(\hat{\theta}_1^{(C)} - \theta_1^*) \xrightarrow{D} N_{q_1}(\mathbf{0}, \Psi_1)$ .

Now,  $\tau_1^{-1/2}(\hat{\tau}_1^{(C)} - \tau_1)$  has also an asymptotic normal distribution because in addition that  $\hat{\theta}_1^{(C)}$  satisfies conditions (i) and (ii) of Theorem 2,  $\hat{\tau}_1^{(C)}$  satisfies condition (iii). Thus by Theorem 2,  $\tau_1^{-1/2}(\hat{\tau}_1^{(C)} - \tau_1) \xrightarrow{D} N(0, \sigma_{1C}^2)$ , where  $\sigma_{1C}^2$  is given by (20).

It is worth noting that the asymptotic marginal distributions of  $\tau_1^{-1/2}(\hat{\tau}_1^{(C)} - \tau_1)$  and  $\tau_1^{1/2}(\hat{\theta}_1^{(C)} - \theta_1^*)$  are not the same as those of  $\tau_1^{-1/2}(\hat{\tau}_1^{(U)} - \tau_1)$  and  $\tau_1^{1/2}(\hat{\theta}_1^{(U)} - \theta_1^*)$ . To show this, we will firstly prove that

$$\Psi_1^{-1} = \Sigma_{122}^{-1} - \frac{1 - n/N}{\pi_0^{(1)}(\theta_1^*) \left[1 - \pi_0^{(1)}(\theta_1^*)\right]} \left[\nabla \pi_0^{(1)}(\theta_1^*)\right] \left[\nabla \pi_0^{(1)}(\theta_1^*)\right]', \quad (37)$$

where  $\Psi_1^{-1}$  is the  $q_1 \times q_1$  matrix defined in the statement of Theorem 2 and  $\Sigma_{122}^{-1}$  is the  $q_1 \times q_1$  submatrix of the matrix  $\Sigma_1^{-1}$ , defined in the statement of Theorem 1, obtained by removing its first row and first column. Since  $\tilde{\pi}_x^{(1)}(\theta_1^*) = \pi_x^{(1)}(\theta_1^*) / \left[1 - \pi_0^{(1)}(\theta_1^*)\right]$ , it follows that

$$\begin{aligned} \frac{\partial \tilde{\pi}_x^{(1)}(\theta_1^*)}{\partial \theta_j^{(1)}} &= \frac{\left[\partial \pi_x^{(1)}(\theta_1^*) / \partial \theta_j^{(1)}\right] \left[1 - \pi_0^{(1)}(\theta_1^*)\right] + \pi_x^{(1)}(\theta_1^*) \left[\partial \pi_0^{(1)}(\theta_1^*) / \partial \theta_j^{(1)}\right]}{\left[1 - \pi_0^{(1)}(\theta_1^*)\right]^2} \\ &= \frac{1}{1 - \pi_0^{(1)}(\theta_1^*)} \frac{\partial \pi_x^{(1)}(\theta_1^*)}{\partial \theta_j^{(1)}} + \frac{\pi_x^{(1)}(\theta_1^*)}{\left[1 - \pi_0^{(1)}(\theta_1^*)\right]^2} \frac{\partial \pi_0^{(1)}(\theta_1^*)}{\partial \theta_j^{(1)}}. \end{aligned}$$

Then

$$\begin{aligned}
& (1 - n/N) \left[ 1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right] \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_i^{(1)}} \frac{\partial \tilde{\pi}_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \\
&= (1 - n/N) \left[ 1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right] \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)} \left[ \frac{1}{1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_i^{(1)}} \right. \\
&\quad \left. + \frac{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\left[ 1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right]^2} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_i^{(1)}} \right] \left[ \frac{1}{1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} + \frac{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\left[ 1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right]^2} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \right] \\
&= (1 - n/N) \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{1}{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_i^{(1)}} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} + \frac{1 - n/N}{1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \\
&\quad \times \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_i^{(1)}} + \frac{1 - n/N}{1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_i^{(1)}} \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} + \frac{1 - n/N}{\left[ 1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right]^2} \\
&\quad \times \frac{\partial \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_i^{(1)}} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*) \\
&= (1 - n/N) \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{1}{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_i^{(1)}} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} - \frac{2(1 - n/N)}{1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_i^{(1)}} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \\
&\quad + \frac{1 - n/N}{1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_i^{(1)}} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \\
&= (1 - n/N) \sum_{\mathbf{x} \in \Omega - \{\mathbf{0}\}} \frac{1}{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_i^{(1)}} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} - \frac{1 - n/N}{1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_i^{(1)}} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} \\
&= (1 - n/N) \sum_{\mathbf{x} \in \Omega} \frac{1}{\pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_i^{(1)}} \frac{\partial \pi_{\mathbf{x}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}} - \frac{1 - n/N}{\pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \left[ 1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right]} \\
&\quad \times \frac{\partial \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_i^{(1)}} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}}.
\end{aligned}$$

Therefore, from the definitions of  $\Psi_1^{-1}$  and  $\Sigma_{122}^{-1}$  we have that

$$\left[ \Psi_1^{-1} \right]_{i,j} = \left[ \Sigma_{122}^{-1} \right]_{i,j} - \frac{1 - n/N}{\pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \left[ 1 - \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*) \right]} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_i^{(1)}} \frac{\partial \pi_{\mathbf{0}}^{(1)}(\boldsymbol{\theta}_1^*)}{\partial \theta_j^{(1)}},$$

and (37) is proved.



From (34) and (37) it follows that  $\sigma_{1C}^2 \neq \sigma_{1U}^2$ , and hence  $\tau_1^{-1/2}(\hat{\tau}_1^{(U)} - \tau_1)$  and  $\tau_1^{-1/2}(\hat{\tau}_1^{(C)} - \tau_1)$  do not have the same asymptotic normal distribution. In addition, (35) and (37) imply that  $\Psi_1 \neq \Sigma_{122}$ , and consequently that  $\tau_1^{1/2}(\hat{\theta}_1^{(U)} - \theta_1^*)$  and  $\tau_1^{1/2}(\hat{\theta}_1^{(C)} - \theta_1^*)$  do not have the same asymptotic normal distribution. Notice also that even though the asymptotic marginal distributions of the UMLEs and CMLEs of  $\tau_1$  and  $\theta_1^*$  are not the same, from (37) it follows that if  $n/N$  were small enough so that  $1 - (1 - n/N)\pi_0^{(1)}(\theta_1^*) \approx 1 - \pi_0^{(1)}(\theta_1^*)$ , then  $\Psi_1 \approx \Sigma_{122}$  and their asymptotic marginal distributions would be very similar to each other.

#### 4.6 Asymptotic properties of unconditional and conditional maximum likelihood estimators of $(\tau_2, \theta_2^*)$

The unconditional and conditional maximum likelihood estimators of  $(\tau_2, \theta_2^*)$  are exactly the same as those used in capture-recapture studies. Sanathanan (1972) assumed conditions similar to (1)-(4) and (6) and proved the following results:

- (i)  $\hat{\theta}_2^{(U)} \xrightarrow{P} \theta_2^*$  and  $\hat{\theta}_2^{(C)} \xrightarrow{P} \theta_2^*$  as  $\tau_2 \rightarrow \infty$ .
- (ii)  $\hat{\tau}_2^{(U)}/\tau_2 \xrightarrow{P} 1$  and  $\hat{\tau}_2^{(C)}/\tau_2 \xrightarrow{P} 1$  as  $\tau_2 \rightarrow \infty$ .
- (iii)  $\left[ \tau_2^{-1/2}(\hat{\tau}_2^{(U)} - \tau_2), \tau_2^{1/2}(\hat{\theta}_2^{(U)} - \theta_2^*) \right] \xrightarrow{D} N_{q_2}(\mathbf{0}, \Sigma_2)$  and  $\left[ \tau_2^{-1/2}(\hat{\tau}_2^{(C)} - \tau_2), \tau_2^{1/2}(\hat{\theta}_2^{(C)} - \theta_2^*) \right] \xrightarrow{D} N_{q_2}(\mathbf{0}, \Sigma_2)$  as  $\tau_2 \rightarrow \infty$ ,

where  $\Sigma_2$  is the inverse of the  $(q_2 + 1) \times (q_2 + 1)$  matrix  $\Sigma_2^{-1}$  defined by

$$\begin{aligned} [\Sigma_2^{-1}]_{1,1} &= [1 - \pi_0^{(2)}(\theta_2^*)] / \pi_0^{(2)}(\theta_2^*), \\ [\Sigma_2^{-1}]_{1,j+1} &= [\Sigma_2^{-1}]_{j+1,1} = -[1/\pi_0^{(2)}(\theta_2^*)] \left[ \partial \pi_0^{(2)}(\theta_2^*) / \partial \theta_j^{(2)} \right], \quad j = 1, \dots, q_2, \\ [\Sigma_2^{-1}]_{i+1,j+1} &= [\Sigma_2^{-1}]_{j+1,i+1} = \sum_{\mathbf{x} \in \Omega} [1/\pi_{\mathbf{x}}^{(2)}(\theta_2^*)] \left[ \partial \pi_{\mathbf{x}}^{(2)}(\theta_2^*) / \partial \theta_i^{(2)} \right] \left[ \partial \pi_{\mathbf{x}}^{(2)}(\theta_2^*) / \partial \theta_j^{(2)} \right], \\ &\quad i, j = 1, \dots, q_2, \end{aligned}$$

and which is assumed to be a non-singular matrix.

Because the proofs of these results are exactly the same as those given by Sanathanan (1972), we will omit them. It is worth noting that unlike the CMLE  $(\tau_1^{-1/2}\hat{\tau}_1^{(C)}, \tau_1^{1/2}\hat{\theta}_1^{(C)})$ , the estimator  $(\tau_2^{-1/2}\hat{\tau}_2^{(C)}, \tau_2^{1/2}\hat{\theta}_2^{(C)})$  does have an asymptotic multivariate normal distribution.

The previous results imply that  $\tau_2^{-1/2}(\hat{\tau}_2^{(U)} - \tau_2) \xrightarrow{D} N(0, \sigma_2^2)$  and  $\tau_2^{-1/2}(\hat{\tau}_2^{(C)} - \tau_2) \xrightarrow{D} N(0, \sigma_2^2)$ , where

$$\sigma_2^2 = \frac{1}{1 - \pi_0^{(2)}(\theta_2^*)} \left\{ \pi_0^{(2)}(\theta_2^*) + \frac{1}{1 - \pi_0^{(2)}(\theta_2^*)} \left[ \nabla \pi_0^{(2)}(\theta_2^*) \right]' \left[ \Sigma_{222}^{-1} - \frac{1}{\pi_0^{(2)}(\theta_2^*) \left[ 1 - \pi_0^{(2)}(\theta_2^*) \right]} \left[ \nabla \pi_0^{(2)}(\theta_2^*) \right] \left[ \nabla \pi_0^{(2)}(\theta_2^*) \right]' \right]^{-1} \left[ \nabla \pi_0^{(2)}(\theta_2^*) \right] \right\},$$

where  $\nabla \pi_0^{(2)}(\theta_2^*)$  is the gradient of  $\pi_0^{(2)}(\theta_2)$  evaluated at  $\theta_2^*$  and  $\Sigma_{222}^{-1}$  is the  $q_2 \times q_2$  submatrix of  $\Sigma_2^{-1}$  obtained by removing its first row and first column.

#### 4.7 Consistency and asymptotic normality of the unconditional and conditional maximum likelihood estimators of $\tau = \tau_1 + \tau_2$

The UMLE and CMLE of  $\tau = \tau_1 + \tau_2$  were defined in Subsection 3.3.3 by  $\hat{\tau}^{(U)} = \hat{\tau}_1^{(U)} + \hat{\tau}_2^{(U)}$  and  $\hat{\tau}^{(C)} = \hat{\tau}_1^{(C)} + \hat{\tau}_2^{(C)}$ . From assumptions **A** and **B** and the previous results we have that  $\hat{\tau}^{(U)}/\tau = (\tau_1/\tau) (\hat{\tau}_1^{(U)}/\tau_1) + (\tau_2/\tau) (\hat{\tau}_2^{(U)}/\tau_2) \xrightarrow{P} \alpha_1 \times 1 + \alpha_2 \times 1 = 1$ , as  $\tau_1 \rightarrow \infty$  and  $\tau_2 \rightarrow \infty$ . Similarly,  $\hat{\tau}^{(C)}/\tau \xrightarrow{P} 1$  as  $\tau_1 \rightarrow \infty$  and  $\tau_2 \rightarrow \infty$ . Furthermore,  $\tau^{-1/2}(\hat{\tau}^{(U)} - \tau) = (\tau/\tau_1)^{-1/2} \tau_1^{-1/2} (\hat{\tau}_1^{(U)} - \tau_1) + (\tau/\tau_2)^{-1/2} \tau_2^{-1/2} (\hat{\tau}_2^{(U)} - \tau_2) \xrightarrow{D} N(0, \sigma_U^2)$ , as  $\tau_1 \rightarrow \infty$  and  $\tau_2 \rightarrow \infty$ , where  $\sigma_U^2 = \alpha_1 \sigma_{1U}^2 + \alpha_2 \sigma_2^2$ . Likewise,  $\tau^{-1/2}(\hat{\tau}^{(C)} - \tau) \xrightarrow{D} N(0, \sigma_C^2)$ , where  $\sigma_C^2 = \alpha_1 \sigma_{1C}^2 + \alpha_2 \sigma_2^2$ .

### 5 Estimation of the matrices $\Sigma_k^{-1}$ and $\Psi_1^{-1}$

Although estimates of  $\Sigma_k^{-1}$ ,  $k = 1, 2$ , and  $\Psi_1^{-1}$  can be obtained by replacing the parameters  $\theta_k^*$  and  $\theta_1^*$  by their respective estimates in the expressions for these matrices, this procedure requires the computation of sums of  $2^n$  terms. This is not a problem if  $n$  is small, but if  $n$  is large enough, say greater than or equal to 20, the number of these terms is very large and the calculation of the estimates of  $\Sigma_k^{-1}$  and  $\Psi_1^{-1}$  could be computationally expensive.

A procedure that requires a much smaller number of calculations is based on estimates of the vectors  $\mathbf{V}_t^{(k)}$ ,  $t = 1, \dots, \tau_k$ ,  $k = 1, 2$ . Vectors  $\mathbf{V}_t^{(1)}$ s were defined in the proofs of Theorems 1 and 2, whereas vectors  $\mathbf{V}_t^{(2)}$ s are defined in Sanathanan (1972) and we will give their definition later in this section. As was shown in the proofs of Theorem 1 and 2, the vectors  $\mathbf{V}_t^{(1)}$ s are independent and equally distributed with mean vector equal to the vector zero and covariance matrix equal to  $\Sigma_1^{-1}$  in the case of Theorem 1, and  $\Psi_1^{-1}$  in the case of Theorem 2. The same result holds in the case of the vectors  $\mathbf{V}_t^{(2)}$ s, but the covariance matrix is  $\Sigma_2^{-1}$ . Therefore, the sample covariance matrix of the vectors  $\mathbf{V}_t^{(k)}$ s is an estimate of their covariance matrix ( $\Sigma_k^{-1}$  or  $\Psi_1^{-1}$ ) based only on  $\tau_k$  observations.

To implement this procedure we need to estimate the  $\mathbf{V}_t^{(k)}$ s (they are unknown because depend on  $\theta_k^*$  and  $\tau_k$ ). In the case of the  $\mathbf{V}_t^{(1)}$ s defined in Theorem 1 an estimate  $\hat{\mathbf{V}}_t^{(1)}$  of  $\mathbf{V}_t^{(1)}$  could be obtained by replacing  $\theta_1^*$  by  $\hat{\theta}_1^{(U)}$  in the expression for  $\mathbf{V}_t^{(1)}$ , and  $\tau_1$  could be estimated by  $\hat{\tau}_1^{(U)}$ . In the case of the  $\mathbf{V}_t^{(1)}$ s defined in Theorem 2 estimates of  $\mathbf{V}_t^{(1)}$  could be obtained by replacing  $\theta_1^*$  by  $\hat{\theta}_1^{(C)}$  in the expression for  $\mathbf{V}_t^{(1)}$ , and  $\tau_1$  could be estimated by  $\hat{\tau}_1^{(C)}$ . Estimates of  $\mathbf{V}_t^{(2)}$ s and  $\tau_2$  could be obtained as in the case of Theorem 1, and in this situation both UMLE and CMLE could be used. Thus, once  $\hat{\tau}_k$  and the vectors  $\hat{\mathbf{V}}_t^{(k)}$ s are obtained, their sample covariance matrix can be computed and used as an estimate of  $\Sigma_k^{-1}$  or  $\Psi_1^{-1}$ .

The vectors  $\mathbf{V}_t^{(2)} = [V_{t,1}^{(2)}, \dots, V_{t,q_2+1}^{(2)}]'$ ,  $t = 1, \dots, \tau_2$ , are defined as follows:

- (a)  $V_{t,1}^{(2)} = 1$  and  $V_{t,j+1}^{(2)} = [\pi_{\mathbf{x}}^{(2)}(\theta_2^*)]^{-1} \partial \pi_{\mathbf{x}}^{(2)}(\theta_2^*) / \partial \theta_j^{(2)}$ ,  $j = 1, \dots, q_2$ , if the vector  $X_t^{(2)}$  of link-indicator variables associated with the  $t$ -th element in  $U_2$  equals the vector  $\mathbf{x} \in \Omega - \{\mathbf{0}\}$ ;
- (b)  $V_{t,1}^{(2)} = -[1 - \pi_0^{(2)}(\theta_2^*)] / \pi_0^{(2)}(\theta_2^*)$  and  $V_{t,j+1}^{(2)} = [\pi_0^{(2)}(\theta_2^*)]^{-1} \partial \pi_0^{(2)}(\theta_2^*) / \partial \theta_j^{(2)}$ ,  $j = 1, \dots, q_2$ , if the vector  $X_t^{(2)}$  of link-indicator variables associated with the  $t$ -th element in  $U_2$  equals the vector  $\mathbf{0} \in \Omega$ .

## 6 Conclusions

Whenever we want to apply the results that we have obtained in this research to an actual situation we need to determine whether or not the assumed conditions are reasonably well satisfied by those observed in the actual scenario. In particular we have assumed that the numbers  $M_i$ s of people found in the sampled sites follow a multinomial distribution with homogeneous cell probabilities and that the  $M_i$ s go to infinity while the number of sites  $n$  in the sample and  $N$  in the frame are fixed. These assumptions imply that in the actual scenario the  $M_i$ s should be relatively large and not very variable. However, we do not know how large they should be so that the results can be safely used. Therefore, Monte Carlo studies are required to assess the reliability of the asymptotic results under different scenarios with finite samples and populations. In addition, although we have assumed a general parametric model for the link-probabilities which allows the possibility that the parameter depends or not on the sampled sites, the model precludes that the probabilities depend on the  $M_i$ s as they go to infinity. Furthermore, this assumption assures that the estimators of  $\tau_1$  and  $\tau_2$  be independent and not only conditionally independent given the  $M_i$ s.

An alternative asymptotic framework to the one considered in this work is to assume that the numbers of sites  $n$  in the sample and  $N$  in the frame go to infinity whereas the  $M_i$ s are fixed. However, this would involve dealing with multinomial distributions with infinite numbers of cells. An approach that could be used to derive asymptotic properties of estimators under this framework is the one considered by Rao (1958) who derived asymptotic properties of a maximum likelihood estimator of a parameter on which depend the cell probabilities of

a multinomial distribution with infinite number of cells. However, this is a topic of a future research.

## Acknowledgements

This research was partially supported by Grant PIFI-2013-25-73-1.4.3-8 from the Secretaría de Educación Pública to Universidad Autónoma de Sinaloa.

## References

- Agresti, A. 2002. *Categorical Data Analysis, Second edition*. New York: Wiley.
- Birch, M.W. 1964. "A new proof of the Pearson-Fisher theorem." *Annals of Mathematical Statistics* 35:718–824.
- Bishop, Y.M.M., S.E. Fienberg, and P.W. Holland. 1975. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press.
- Coull, B. A., and A. Agresti. 1999. "The use of mixed logit models to reflect heterogeneity in capture-recapture studies." *Biometrics* 55:294–301.
- Félix-Medina, M.H., and P.E. Monjardin. 2006. "Combining link-tracing sampling and cluster sampling to estimate the size of hidden populations: a Bayesian assisted approach." *Survey Methodology* 32:187–195.
- Félix-Medina, M.H., P.E. Monjardin, and A.N. Aceves-Castro. 2015. "Combining link-tracing sampling and cluster sampling to estimate the size of a hidden population in presence of heterogeneous link-probabilities." *Survey Methodology*. To appear.
- Félix-Medina, M.H., and S.K. Thompson. 2004. "Combining cluster sampling and link-tracing sampling to estimate the size of hidden populations." *Journal of Official Statistics* 20:19–38.
- Feller, W. 1968. *An Introduction to Probability Theory and its Applications, Third edition*. Volume 1. New York: Wiley.
- Johnston, L.G., and K. Sabin. 2010. "Sampling hard-to-reach populations with respondent driven sampling." *Methodological Innovations Online* 5 (2): 38–48.
- Kalton, G. 2009. "Methods for oversampling rare populations in social surveys." *Survey Methodology* 35:125–141.
- Magnani, R., K. Sabin, T. Saidel, and D. Heckathorn. 2005. "Review of sampling hard-to-reach populations for HIV surveillance." *AIDS* 19:S67–S72.
- Rao, C.R. 1958. "Maximum likelihood estimation for the multinomial distribution with infinite number of cells." *Sankhyā: The Indian Journal of Statistics* 20:211–218.
- . 1973. *Linear Statistical Inference and its Applications, Second edition*. New York: Wiley.

- Sanathanan, L. 1972. "Estimating the size of a multinomial population." *Annals of Mathematical Statistics* 43:142–152.
- Spreen, M. 1992. "Rare populations, hidden populations and link-tracing designs: what and why?" *Bulletin de Méthodologie Sociologique* 36:34–58.
- Thompson, S.K., and O. Frank. 2000. "Model-based estimation with link-tracing sampling designs." *Survey Methodology* 26:87–98.
- Varadhan, S.R.S. 2008. "Large deviations." *The Annals of Probability* 2:397–419.